Journal of Experimental and Theoretical Physics, Vol. 93, No. 4, 2001, pp. 671–676.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 771–777.

GRAVITATION, ASTROPHYSICS

Spiral Magneto-Electron Waves in Interstellar Gas Dynamics¶

S. I. Bastrukov[a, b, c, *], J. Yang [b, c], and D. V. Podgainy[a]

[a]*Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia*
[b]*Center for High Energy Astrophysics, Ewha Womans University, Seoul, 120-750, Korea*
[c]*Center for High Energy Physics, Kyungpook National University, 702-701, Daegu, Korea*
**e-mail: bast@thsun1.jinr.ru*

Received March 29, 2001

Abstract—We discuss possible observational consequences resulting from the propagation of transverse magneto-electron waves in the interstellar medium. We briefly describe a magnetohydrodynamic model for the cyclotron waves with emphasis on their analogy with hydrodynamic inertial waves. It is shown that the cyclotron waves are heavily damped in the interstellar medium and, therefore, cannot affect the gas dynamics of star-forming molecular clouds. We developed an analytical model of the helicoidal magneto-electron waves based on the electromagnetic induction equation for the magnetic flux density driven by the Hall and Ohmic components of the electric field generated by flows of thermal electrons. It is established that the helicons can propagate in the interstellar medium without any noticeable attenuation. The presented numerical estimates for the group velocity of the intercloud helicons suggest that spiral circularly polarized magneto-electron waves of this type can be responsible for the broadening of molecular lines detected from dark interstellar clouds. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The gasdynamic processes in star-forming molecular clouds are primarily determined by a strong coupling of the gas–dust interstellar medium (ISM) to magnetic fields. The electrons, both relativistic and thermal, are one of the most abundant and mobile charged components of the ISM, and their small mass provides the most strong coupling to the intercloud magnetic fields. Therefore, it is natural to expect that the collective behavior of electrons may essentially affect interstellar gas dynamics.

According to the available data on the pulsar dispersion measure, the average density of interstellar electrons evaluated throughout the Galactic disc is estimated as $n_e \approx 0.03$ cm^{-3} [1]. On the one hand, an extensive analysis of the thermodynamic state of interstellar medium in our Galaxy reported by Heiles and Kulkarni [2] implies that the warm diffusive clouds might be the regions of the most dense accommodation of thermal electrons where their density can attain a sufficiently high value $n_e \approx 1$ cm^{-3} [3]. On the other hand, highly ionized HII regions of the warm interstellar medium occupy only 25% of the Galactic volume [4]. Therefore, they give a significant contribution to the dispersion measure for only a small fraction of pulsars. The latter observation was made long ago by Manchester and Taylor [5] and led them to suggest that dispersing electrons can basically be located in denser interstellar clouds highly obscured for the ionizing ultraviolet radiation and soft X-rays. However, if the thermal electrons reside in the central region of a molecular cloud at some stage of the star formation, their detection becomes a highly formidable problem. The progress in searching for their presence could then be achieved by inspecting observational consequences caused by highly coherent electron gasdynamical processes.

This paper discusses two wave processes associated with collective motions of interstellar electrons in the presence of a uniform magnetic field. The first process represents the well-known cyclotron waves originating from the inertial collective motions of electrons driven by the Lorentz force. These waves are briefly discussed in Section 2. Emphasis in this paper is placed on the second wave process, representing the helicoidal magneto-electron waves inherently related to the Hall drift of the magnetic flux density by the flows of thermal electrons. The model of interstellar helicons is developed in Section 3. For a deeper insight into the physics of these spiral circularly polarized waves, we confine our consideration to the idealized model of a gaseous magnetically supported cloud whose gas dynamics is dominated by thermal electrons in the regime of strong coupling between the densities of the electron current and the magnetic flux threading the cloud; the mobility of ions and neutral molecules is assumed to be heavily suppressed. In doing this, we clearly realize that ignoring the dusty component of the intercloud medium and assuming immobility of ions and neutrals, thereby eliminating the ambipolar diffusion effect, can be a

¶This article was submitted by the authors in English.

matter of controversy. Nevertheless this does not affect the interest in the problem of the wave gas dynamics of interstellar electrons as such. To the best of our knowledge, the problem of the wave transport of magnetic field by thermal electrons in the interstellar medium has not been considered in the literature. In Section 4, based on the electromagnetic induction equation for the transport of the magnetic flux density by the Hall and Ohmic components of the electric field generated by flows of thermal electrons, we derive the dispersion relation for helicons and evaluate their group velocity. We show that the helicoidal magneto-electron waves can propagate in the interstellar medium without a noticeable attenuation. This allows us to suggest that the helicons could be responsible for the observed widths of molecular lines detected from dark interstellar clouds. In Section 5, a brief outlook is given of the wave process considered here and another wave process in dark interstellar clouds affecting the broadening of molecular lines.

2. INTERSTELLAR CYCLOTRON WAVES

The model of intercloud cyclotron waves can be developed on the basis of hydrodynamical equations for the collective motions of electrons in a permanent magnetic field **B**. Modeling the interstellar electron gas as a viscous uniformly charged fluid whose flows are governed by the Lorentz force, we can write

$$\frac{d\rho_e}{dt} + \rho_e \nabla \cdot \mathbf{u} = 0, \quad \frac{d}{dt} = \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla, \tag{1}$$

$$\rho_e \frac{d\mathbf{u}}{dt} = \frac{en_e}{c}\mathbf{u} \times \mathbf{B} + \eta \Delta \mathbf{u}, \quad \rho_e = n_e m_e, \tag{2}$$

where m_e is the mass of the electron, n_e is the electron density, **u** is the directed velocity of the electron flow, and η stands for the dynamical viscosity of the electron fluid effectively accounting for all the dissipative effects of elastic collisions of electrons with other microparticles of the cloud. Viewing the electrons as an incompressible fluid, we linearize the above equations around the homogeneous undisturbed state with $\mathbf{u}_0 = 0$ and $\mathbf{B}_0 = B\mathbf{e}_z$. As a result, we obtain [1]

$$\nabla \cdot \delta\mathbf{u} = 0, \quad \frac{\nabla \delta\mathbf{u}}{\partial t} = \frac{e}{m_e c}\delta\mathbf{u} \times \mathbf{B} + \frac{\eta}{\rho_e}\Delta\delta\mathbf{u}. \tag{3}$$

Using the plane-wave form for

$$\delta\mathbf{u} \approx \exp\left[i(\mathbf{k} \cdot \mathbf{r} - \omega t)\right],$$

in the first of Eqs. (3), we have

$$\mathbf{k} \cdot \delta\mathbf{u} = 0,$$

which implies that the wave is transverse. Inserting the plane-wave form of $\delta\mathbf{u}$ into equation of motions in (3), after some algebra we obtain the dispersion relation

$$\omega = \pm\omega_c(1 \pm i\Gamma_c), \quad \omega_c = \frac{eB}{m_e c}, \quad \Gamma_c = \frac{\eta k^2}{\omega_c \rho_e}. \tag{4}$$

It then follows that the interstellar cyclotron waves are transverse, circularly polarized, and damped. Their group velocity is given by

$$\mathbf{V}_c = \frac{\partial\omega}{\partial\mathbf{k}} = \pm\frac{e}{m_e c}\frac{\mathbf{k} \times [\mathbf{B} \times \mathbf{k}]}{k^3}. \tag{5}$$

To describe the effect of cyclotron waves on the molecular linewidths, we consider the case where $\mathbf{k} \perp \mathbf{B}$. Recalling the mean electron density $n_e = 10^{-3}$ cm^{-3}, the average magnetic field intensity $B \sim 10^{-5}$ G, and the average electron temperature T_e = 10–200 K in dark molecular clouds [7] and putting $V_c \approx 0.3$–0.5 km/s (the realm of observed molecular linewidths), we can evaluate the wavelength of a cyclotron wave as $\lambda_c \approx 10^2$–10^3 cm. The viscosity coefficient in the interstellar electron fluid can be evaluated as [8]

$$\eta \approx \frac{n_e k_B Te}{\nu_c} \text{ g/cm s.}$$

The frequency of the elastic collisions of electrons is typically in the interval $\nu_c \sim 10^{-3}$–10^{-1} s^{-1} [7]. With the above parameters, the magnitude of the damping coefficient $\Gamma_c \approx 10^3$–10^4 is in strong conflict with the propagation criterion $\Gamma_c \ll 1$ following from Eq. (4). The same can be said about diffusive clouds. Thus, the cyclotron waves of the Larmor gyration of the electron flow about the direction of the equilibrium magnetic field threading the cloud are highly damped. Therefore, they cannot produce any essential effect on molecular linewidths detected from interstellar molecular clouds.

3. MAGNETIC FLUX EVOLUTION DUE TO THE HALL AND OHMIC CONDUCTIVITIES OF THERMAL ELECTRONS

The cyclotron waves characterize a high-frequency branch of the collective oscillatory behavior of electrons in the interstellar magnetoplasma. In this section, we focus on the low-frequency magneto-electron waves whose origin is in the transport of the magnetic flux density by thermal electrons, or in other words, in the Hall electron conductivity. It should be mentioned that the magneto-electron waves under consideration were first discovered in the solid-state plasma physics. The name helicons [9–11] was coined because of the spiral character of these circularly polarized waves. By saying that the helicons represent a low-frequency branch of electromagnetic excitations in a noncompensated electron-dominated magnetoplasma, we imply that the frequency of electron oscillations in this wave

[1] It is worth noting that cyclotron waves can be considered as an analog of inertial waves in the rotating incompressible fluid governed by the equations [6]

$$\nabla \cdot \delta\mathbf{u} = 0, \quad \frac{\partial\delta\mathbf{u}}{\partial t} = 2\delta\mathbf{u} \times \boldsymbol{\Omega} + \nu\Delta\delta\mathbf{u},$$

where $\nu = \eta/\rho$ is the kinematic viscosity.

is less than the cyclotron frequency (see, e.g., [12]). Waves of a similar nature are observed in planetary magnetospheres. In particular, the propagation of helicons in the Earth's ionosphere causes the whistling audio noise on radio, which is why these waves are often called whistlers [12, 13]. A similar Hall mechanism of the wave transport of a magnetic flux density by electrons was recently discussed in [14, 15] in the context of the magnetic field evolution in radio pulsars and magnetars.

In what follows, we confine our consideration to the idealized model of the isothermal intercloud medium whose gas dynamics is dominated by thermal electrons. The magnetic field is considered to be frozen into the ions, the mobility of the ions and of the neutral molecules is assumed to be heavily suppressed, and the latter are therefore regarded as immobilized. This suggests that the collective behavior of intercloud electrons in the presence of a permanent magnetic field has some common features with the behavior of conducting electrons in a metal solid, where the immobility of ions is taken for granted. Following this line of arguments, we use of the constitutive equation for the electron conductivity in the form of the generalized Ohm law

$$\begin{aligned}\mathbf{E}(\mathbf{r}, t) &= \frac{\mathbf{j}(\mathbf{r}, t)}{\sigma_C} + \frac{1}{ecn_e}\mathbf{B}(\mathbf{r}, t)\times\mathbf{j}(\mathbf{r}, t)\\ &= \frac{\mathbf{j}}{\sigma_C} + \frac{\mathbf{n}_C\times\mathbf{j}}{\sigma_H}, \quad \mathbf{n}_B = \frac{\mathbf{B}}{B},\end{aligned} \tag{6}$$

$$\sigma_C = \frac{n_e e^2}{m_e \nu_c}, \quad \sigma_H = \frac{en_e c}{B}, \tag{7}$$

where the Ohmic conductivity σ_C is given by the Drude formula and σ_H stands for the Hall conductivity. We let $\mathbf{j}(\mathbf{r}, t)$ denote the electron current density that is given by

$$\mathbf{j}(\mathbf{r}, t) = \frac{c}{4\pi}\nabla\times\mathbf{B}(\mathbf{r}, t) \tag{8}$$

in accordance with the Ampére law. Because the magnetic flux density satisfies the Maxwell equation for the Faraday induction,

$$\frac{\partial\mathbf{B}(\mathbf{r}, t)}{\partial t} = -c\nabla\times\mathbf{E}(\mathbf{r}, t), \tag{9}$$

inserting Eqs. (6) and (8) in (9) gives

$$\begin{aligned}\frac{\partial\mathbf{B}(\mathbf{r}, t)}{\partial t} &= -\frac{c}{4\pi en_e}\nabla\times[\mathbf{B}(\mathbf{r}, t)\times\nabla\times\mathbf{B}(\mathbf{r}, t)]\\ &\quad + \frac{c^2}{4\pi\sigma_C}\Delta\mathbf{B}(\mathbf{r}, t).\end{aligned} \tag{10}$$

We note that this model can be considered as an idealized version of the model motivated by Mouschovias [7] in the context of the magnetic flux redistribution in cores of the interstellar magnetically supported clouds with the ions frozen into the magnetic field threading the cloud. The first term in the right-hand side of Eq. (10) is due to the Hall electron conductivity and the second term describes the Ohmic diffusion of the magnetic field. Obviously, the diffusion-free regime of the magnetic flux transport is realized if

$$\sigma_C \gg \sigma_H. \tag{11}$$

In this regime, Eq. (10) supplemented by the sole noidality condition for $\mathbf{B}(\mathbf{r}, t)$ becomes

$$\begin{aligned}\frac{\partial\mathbf{B}(\mathbf{r}, t)}{\partial t} &= -\frac{c}{4\pi en_e}\nabla\times[\mathbf{B}(\mathbf{r}, t)\times[\nabla\times\mathbf{B}(\mathbf{r}, t)]],\\ \nabla\cdot\delta\mathbf{B}(\mathbf{r}, t) &= 0,\end{aligned} \tag{12}$$

which implies that the total magnetic energy

$$W_m = \frac{1}{8\pi}\int\mathbf{B}^2 dV \tag{13}$$

is conserved:

$$\begin{aligned}\frac{dW_m}{dt} &= \frac{1}{4\pi}\int\mathbf{B}\cdot\frac{\partial\mathbf{B}}{\partial t}dV = \frac{c}{16\pi^2 en_e}\\ &\times\int[[\nabla\times\mathbf{B}]\times\mathbf{B}]\cdot[\nabla\times\mathbf{B}]dV = 0.\end{aligned} \tag{14}$$

It is the major purpose of the remainder of this paper to show that the interstellar magnetoplasma of molecular star-forming clouds can transmit low-frequency perturbations in the magnetic flux density by weakly damped helicoidal circularly polarized waves owing their existence to the Hall drift of the magnetic field by flows of thermal electrons.

4. HELICONS IN THE INTERSTELLAR MEDIUM

We consider the evolution of small-amplitude magnetic flux density perturbations $\delta\mathbf{B}$ superimposed on the permanent magnetic field $\mathbf{B}$,

$$\mathbf{B}(\mathbf{r}, t) \longrightarrow \mathbf{B} + \delta\mathbf{B}(\mathbf{r}, t), \quad \mathbf{B} = \text{const}. \tag{15}$$

The corresponding linearization of Eq. (10) leads to

$$\frac{\partial\delta\mathbf{B}}{\partial t} = \frac{c}{4\pi n_e e}(\mathbf{B}\cdot\nabla)[\nabla\times\delta\mathbf{B}] + \frac{c^2}{4\pi\sigma_C}\Delta\delta\mathbf{B}. \tag{16}$$

Inserting

$$\delta\mathbf{B} = \mathbf{b}\exp[i(\omega t - \mathbf{k}\cdot\mathbf{r})], \tag{17}$$

in the right-hand side of Eq. (16), we obtain

$$\begin{aligned}\frac{\partial\delta\mathbf{B}}{\partial t} &= -\frac{c^2}{4\pi\sigma_H}(\mathbf{k}\cdot\mathbf{n}_B)[\mathbf{k}\times\delta\mathbf{B}]\\ &\quad + \frac{c^2}{4\pi\sigma_C}[\mathbf{k}\times[\mathbf{k}\times\delta\mathbf{B}]].\end{aligned} \tag{18}$$

We let the permanent field $\mathbf{B}$ be directed along the z axis, $\mathbf{B} = [0, 0, B]$, and consider a one-dimensional plane-wave perturbation along the z axis ($\mathbf{k} = k\mathbf{e}_z$) that does not affect the intensity of the magnetic field in this

direction, but only along the x an y directions. This means that the fluctuating magnetic field components depend only on z and t:

$$\delta B_x(z,t) = b_x \exp[i(\omega t - kz)],$$
$$\delta B_y(z,t) = b_y \exp[i(\omega t - kz)], \quad \delta B_z = 0. \tag{19}$$

The Cartesian components of Eq. (18) can be represented in the equivalent form

$$\frac{\partial \delta B_x}{\partial t} = +\frac{c^2k^2}{4\pi\sigma_H}\delta B_y - \frac{c^2k^2}{4\pi\sigma_C}\delta B_x, \tag{20}$$

$$\frac{\partial \delta B_y}{\partial t} = -\frac{c^2k^2}{4\pi\sigma_H}\delta B_x - \frac{c^2k^2}{4\pi\sigma_C}\delta B_y. \tag{21}$$

To see the circularly polarized character of the wave motions in question, we omit the Ohmic diffusion term for the moment. The resulting equations (20) and (21) then become

$$\delta \dot{B}_x = -\Omega \delta B_y, \quad \delta \dot{B}_y = \Omega \delta B_x.$$

These are the Cartesian components of the vector equation

$$\delta \dot{\mathbf{B}} = [\Omega \times \delta \mathbf{B}]$$

describing the precession of the vector $\delta\mathbf{B}$ about the z axis with the angular frequency

$$\mathbf{\Omega} = -\frac{cB}{4\pi n_e e}k^2 \mathbf{e}_z.$$

Wave motions of this type are customarily described in terms of the right-hand (δB_+) and the left-hand (δB_-) circularly polarized wave fields

$$\delta B_+ = \delta B_x + i\delta B_y = b(z)\exp(-i\omega t),$$
$$\delta B_- = \delta B_x - i\delta B_y = b(z)\exp(i\omega t). \tag{22}$$

Combining the coupled equations (20) and (21) to obtain one equation for either δB_+ or δB_-, we arrive at

$$\frac{\partial \delta B_\pm}{\partial t} = \mp i\frac{c^2k^2}{4\pi\sigma_H}\delta B_\pm \pm \frac{c^2k^2}{4\pi\sigma_C}\delta B_\pm. \tag{23}$$

Eliminating the time derivative with the help of (22), we obtain

$$\omega = \omega_h(1 \mp i\Gamma_h), \quad \omega_h = \frac{c^2k^2}{4\pi\sigma_H} = \frac{\omega_c}{\omega_p^2}c^2k^2,$$
$$\Gamma_h = \frac{\sigma_H}{\sigma_C} = \frac{\nu_c}{\omega_c}, \tag{24}$$

where

$$\omega_c = \frac{eB}{m_e c}, \quad \omega_p^2 = \frac{4\pi e^2 n_e}{m_e},$$

$$\sigma_C = \frac{n_e e^2}{m\nu_c}, \quad \sigma_H = \frac{en_e c}{B},$$

the omegas stand for the cyclotron and the plasma frequencies, and the sigmas for the Ohmic and Hall conductivities. Dispersion relation (24) implies that the helicon is a transverse circularly polarized and damped wave in which the densities of the magnetic flux and of the electron current undergo coherent oscillations in the plane perpendicular to the direction of propagation. In the diffusion-free regime, $\Gamma_h \ll 1$, we have

$$\omega_h = \frac{\omega_c}{\omega_p^2}c^2k^2 = 4.97\times 10^{18}\frac{B}{n_e}k^2. \tag{25}$$

The corresponding group velocity is given by

$$V_h = \frac{2c^2\omega_c}{\omega_p^2}k \approx 9.58\times 10^{18}\frac{B}{n_e}k \text{ cm/s}. \tag{26}$$

Using (25), we can represent this formula in terms of ω as

$$V_h = \frac{2c\sqrt{\omega_c}}{\omega_p}\sqrt{\omega} \approx 4.46\times 10^9\sqrt{\frac{B}{n_e}}\sqrt{\omega} \text{ cm/s}. \tag{27}$$

In the electron magnetohydrodynamics [16], the helicons play the same role as the transverse Alfvén waves in the single-component magnetohydrodynamics [17]. In both kinds of these magnetohydrodynamic (MHD) waves, the oscillatory motions of the conducting fluid are strongly coupled to the magnetic field fluctuations. The essential kinematic difference between them is that the group velocity of a helicoidal magnetoelectron wave depends on the frequency, whereas the Alfvén wave is characterized by the dispersion-free propagation law

$$\omega = V_A k, \quad V_A = \frac{B}{(4\pi\rho)^{1/2}}.$$

We now briefly discuss inferences that could be made from the propagation of helicoidal magnetoelectron waves in the interstellar medium. As mentioned above, the average density of electrons evaluated in the Galactic disc is estimated to be $n_e \approx 0.03$ cm^{-3} and $B \approx 10^{-5}$–10^{-6} G [1]. With these parameters, we find that the cyclotron frequency ω_c (which sets the upper frequency limit for the dissipation-free propagation of helicons) falls into the interval $10 < \omega_c < 100$ s^{-1}. Taking $\omega \approx 1$ s^{-1} as a representative example, which is typical of pulsar activity, we can estimate the velocity of interstellar helicons as $V_h \approx 10^7$–10^8 cm/s; the corresponding wavelength is $\lambda_h = 2\pi/k \approx 1000$ km. Because the frequency of elastic collisions is of the order 10^{-3} s^{-1} or less, it seems plausible that the helicons can freely travel in the interstellar space and give a contribution to the observed effect of scintillations of the pulsar signals.

The evidence for the existence of large-scale motions in dark star-forming molecular clouds is provided by the widths of molecular lines. Therefore, searching adequate models of interstellar gas dynamics is one of the important parts in the current investigations of the ISM physics. By inspecting a possible effect of helicons on the widths of molecular lines, we note that in a typical dark molecular cloud, $\omega_c \approx 10^2$ s^{-1} and $\nu_c \approx 10^{-3}$–10^{-1} s^{-1}. The criterion for the dissipation-free propagation of helicons $\Gamma = \nu_c/\omega_c \ll 1$ is therefore well justified, and its validity remains quite robust to the changes in ν_c up to $\nu_c = 100$ s^{-1}. By taking the group velocity of the helicons V_h to be equal to the velocity dispersion measured for molecular lines, $V \approx$ 0.3–5.0 km/s, we find the wavelength of the intercloud helicon $\lambda_h \sim 10^{12}$–10^{13} cm. This space scale is much less than the linear size of clouds, $L \sim 10^{17}$ cm. For the same velocity, the period of oscillations of the electron flow in the helicoidal magneto-electron wave falls into the interval $P_h \sim$ 0.1–10 years. These estimates unambiguously show that the intercloud medium can transmit the helicons without significant attenuation and we conjecture that they could be responsible for the broadening of molecular lines.

5. DISCUSSION

Understanding gas dynamical processes governing the structure and the evolution of dense molecular clouds is one of the outstanding challenges in the current development of star formation astrophysics. While the central role played by magnetic fields in these processes was recognized many years ago, the major uncertainties regarding the motions follow from inadequate knowledge of the material composition of the intercloud medium. Over the years, convincing evidence has been obtained that the composition of dark molecular clouds is dominated by molecular hydrogen with some admixture of OH and CO molecules whose linewidths were found to exhibit the supersonic character of intercloud motions. The fact that the linewidths cannot be explained as a result of the propagation of isothermal sound waves has served as an impetus in searching for alternative models of interstellar gas dynamics and has led to the hypothesis that a sizable fraction of charged particles (primarily electrons and ions) are present in dark molecular clouds, with the collective flows of these particles strongly coupled to the intercloud magnetic field. On the assumptions that the magnetic field causes both electrons and ions to move with equal velocities and the friction then causes the neutral molecules to follow the ions with the same velocity, the single-component MHD model was extensively exploited in interpreting supersonic broadening of molecular lines in terms of hydromagnetic waves of the Alfvén type [19–23]. On average, the model provides a reasonable account of data in CO regions of clouds where the temperature and the ionization factor are sufficiently high.

Together with this, recent Zeeman measurements of magnetic fields in dense cores of molecular clouds, highly obscured by the ionizing ultraviolet radiation, have revealed a predominately sub-Alfvénic character of the intercloud motions [24]. The latter circumstance can be regarded as an indication that the composition and the character of the motions in cores of the molecular clouds might be quite different from those that are implied by the single-component MHD model of interstellar gas dynamics. With this in mind, we recently investigated a model of a non-MHD type [25, 26]. Motivated by the observable filamentary structure of some of the dark molecular clouds, we argued that the filaments could be regarded as a manifestation of a superparamagnetic state of the gas–dust ISM considered by Jones and Spitzer [27] long ago in the context of the starlight polarization problem. The magnetically polarized, poorly conducting soft matter of this type can be thought of as a gas-based ferrocolloid (consisting of tiny ferromagnetic grains suspended in the dense gas of molecular hydrogen) capable of sustaining a long-range magnetic chains extending along the intercloud magnetic fields. Having assumed that the motions of the Jones–Spitzer matter are governed by the magnetoelastodynamics equations, we found in [25] that ferrocolloidal interstellar medium can transmit perturbations by shear magnetomechanical waves propagating with a sub-Alfvénic group velocity in accordance with observations [24].

In the meantime, several authors have argued that the ISM motions in star-forming molecular clouds can pass the regime in which the Hall conductivity may become important [28, 29]. In particular, it was recently shown [30] that the Hall conductivity can essentially affect the propagation of Alfvén waves in a dense weakly ionized molecular gas. In this paper, continuing investigation in this direction, we have explored two models of pure electron interstellar gas dynamics. The focus was placed on the helicons (spiral magneto-electron waves owing their existence to the Hall drift of the magnetic flux by thermal electrons). The basic inference of this model is that in dark molecular clouds, the helicons can propagate without a noticeable attenuation. The observational consequence of their propagation might be the widths of molecular lines exhibiting the existence of large-scale intercloud motions. Our numerical estimates for the group velocity of the helicons suggest that these waves could be responsible for the broadening of molecular linewidths detected from dark star-forming clouds or, at least, provide a sizable contribution to this effect. As a conclusion, the problem of the interstellar electron wave dynamics considered here for the first time is interesting in its own right, and we hope that our analysis can find other useful physical applications.

ACKNOWLEDGMENTS

The work was partly supported by most through the Korean National R&D program for women's universities.

REFERENCES

1. A. G. Lyne, R. N. Manchester, and J. H. Taylor, Mon. Not. R. Astron. Soc. **213**, 613 (1985).
2. C. Heiles and S. B. Kulkarni, in *Physical Processes in Interstellar Clouds*, Ed. by E. Morfill and M. Scholer (Reidel, Dordrecht, 1987), p. 13.
3. C. Heiles, L. M. Haffner, R. M. Reynolds, and S. J. Tufte, Astrophys. J. **536**, 335 (2000).
4. C. Heiles, in *Physics of the Gaseous and Stellar Discs of the Galaxy*, Ed. by I. R. King, Astron. Soc. Pac. Conf. Ser. **66**, 249 (1994).
5. R. N. Manchester and J. H. Taylor, *Pulsars* (Freeman, San Francisco, 1977).
6. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics* (Nauka, Moscow, 1986; Pergamon, New York, 1995).
7. T. Ch. Mouschovias, in *Physical Processes in Interstellar Clouds*, Ed. by E. Morfill and M. Scholer (Reidel, Dordrecht, 1987), p. 491.
8. V. E. Golant, A. P. Zhilinskii, and I. E. Sakharov, *Fundamentals of Plasma Physics* (Atomizdat, Moscow, 1977; Wiley, New York, 1980).
9. O. K. Konstantinov and I. V. Perel', Zh. Éksp. Teor. Fiz. **38**, 161 (1960) [Sov. Phys. JETP **11**, 117 (1960)].
10. K. A. Kaner and V. G. Skobov, Usp. Fiz. Nauk **89**, 367 (1966) [Sov. Phys. Usp. **9**, 480 (1967)].
11. P. M. Platzmann and P. A. Wolff, *Waves and Interactions in Solid State Plasmas* (Academic, New York, 1970).
12. A. I. Akhiezer, I. A. Akhiezer, R. V. Polovin, A. G. Sitenko, and K. N. Stepanov, *Plasma Electrodynamics* (Nauka, Moscow, 1974; Pergamon, Oxford, 1975).
13. G. K. Parks, *Physics of Space Plasmas* (Addison-Wesley, London, 1991).
14. P. Goldreich and A. Reisenegger, Astrophys. J. **395**, 250 (1992).
15. C. Thompson and R. C. Duncan, Astrophys. J. **473**, 322 (1996).
16. A. D. Kingsep, K. V. Chukbar, and V. V. Yan'kov, Fiz. Plazmy **16**, 243 (1990) [Sov. J. Plasma Phys. **16**, 136 (1990)].
17. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media* (Nauka, Moscow, 1982; Pergamon, New York, 1995).
18. J. Arons and C. E. Max, Astrophys. J. Lett. **196**, L77 (1975).
19. P. C. Myers and A. Goodman, Astrophys. J. Lett. **326**, L27 (1988).
20. E. G. Zweibel and C. F. McKee, Astrophys. J. **440**, 686 (1995).
21. C. F. Gammie and E. C. Ostriker, Astrophys. J. **466**, 814 (1996).
22. P. Padoan and Å. Nordlund, Astrophys. J. **526**, 279 (1999).
23. W. Pillip, T. W. Hartquist, O. Havnes, and G. E. Morfill, Astrophys. J. **314**, 341 (1987).
24. R. M. Crutcher, Astrophys. J. **520**, 706 (1999).
25. J. Yang and S. I. Bastrukov, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 577 (2000) [JETP Lett. **71**, 395 (2000)].
26. S. I. Bastrukov and J. Yang, Astrophysics **43**, 405 (2000).
27. R. V. Jones and L. Spitzer, Astrophys. J. **146**, 943 (1967).
28. C. Norman and J. Heyvaerts, Astron. Astrophys. **147**, 247 (1985).
29. M. Wardle, Mon. Not. R. Astron. Soc. **298**, 507 (1998).
30. M. Wardle and C. Ng, Mon. Not. R. Astron. Soc. **303**, 249 (1999).

Journal of Experimental and Theoretical Physics, Vol. 93, No. 4, 2001, pp. 677–684.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 778–785.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Domain Walls Carrying a $U(1)$ Charge

V. A. Lensky[a, *], V. A. Gani[b, **], and A. E. Kudryavtsev[b, *]**

[a] *Moscow State Engineering Physics Institute (Technical University), Kashirskoe sh. 31, Moscow, 115409 Russia*
**e-mail: lensky@vxitep.itep.ru*
[b] *State Research Center, Institute of Theoretical and Experimental Physics, Bol'shaya Cheremushkinskaya ul. 25, Moscow, 117259 Russia*
***e-mail: gani@heron.itep.ru*
****e-mail: kudryavt@heron.itep.ru*

Received April 20, 2001

Abstract—A classical field system is considered that consists of two interacting scalar fields, the Higgs real field and a complex scalar field. It is demonstrated that there exists a nontrivial topological solution in this system—a kink carrying a $U(1)$ charge. Certain questions are discussed related to the stability of the solution obtained. An improved variational procedure is proposed for determining topological $U(1)$-charged configurations. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is well known that domain-wall-type solutions arise in a wide class of modern supersymmetric field theories as well as in the string description of gauge theories (D-branes) [1]. The problem of interaction between particles and domain walls has been repeatedly considered in the literature starting from Voloshin's paper [2]. The interaction between abelian gauge particles and domain walls was considered in greater detail in [3]. Usually, the interaction between particles and a wall is considered as scattering in a given potential induced by the wall; i.e., one applies a weak-coupling approximation and neglects the reaction of fields on the wall. Such an approximation cannot always be correct. For example, it is known that the interaction of monopoles and skyrmions with domain walls is nontrivial [4, 5]. Moreover, the walls may form bound states with other fields and have quantum numbers corresponding to these fields. Recently, an integrable dyon-type lattice model with a topological soliton carrying a $U(1)$ charge was discussed in [6].

In the present paper, we consider a simpler continuous model for a system of two interacting scalar fields. In this model, which was discussed long ago [7], there exist nontopological solutions with a $U(1)$ charge, the so-called Q-balls. In this paper, we will show that there exist kink-type topological solutions carrying a $U(1)$ charge in (1 + 1) dimensions in the Lee–Friedberg–Sirlin model.

It should be noted that monopole-type solutions that additionally carry an electric charge (dyons) have been known for a long time in the literature. Other examples of topological solutions carrying a $U(1)$ charge are the so-called Q-lumps [8].

2. TOPOLOGICAL AND NONTOPOLOGICAL Q-BALLS

Consider a system of interacting scalar fields in (1 + 1) dimensions with the Lagrangian

$$\mathcal{L} = \partial_\mu \xi \partial^\mu \xi^* + \frac{1}{2}\partial_\mu \phi \partial^\mu \phi - h^2 \phi^2 \xi \xi^* - \frac{m^2}{2}(\phi^2 - v^2)^2. \quad (1)$$

Here, ϕ is a real scalar field; ξ is a complex scalar field; h, m, and v are real constants; and $\mu = 0, 1$. Such a field Lagrangian in (3 + 1) dimensions was first considered in [7]. In particular, classical spherically symmetric solutions, carrying a $U(1)$ charge, to the equations of motion for Lagrangian (1) were obtained in [7]. Later, these nontopological soliton solutions were referred to as Q-balls [9]. One-dimensional Q-balls for a field system with the Lagrangian close to (1) were discussed in detail in [10].

Lagrangian (1) is symmetric with respect to the global $U(1)$ transformations

$$\xi \longrightarrow \xi e^{i\alpha}, \quad (2)$$

as well as with respect to the discrete transformations

$$\phi \longrightarrow -\phi. \quad (3)$$

The invariance of the Lagrangian under transformations (2) corresponds to the conservation of the current

$$j^\mu = \frac{1}{i}(\xi^* \partial^\mu \xi - \xi \partial^\mu \xi^*)$$

and charge

$$Q = \int j^0 dx.$$

The symmetry (3) corresponds to the conservation of the topological current

$$i^\mu = \epsilon^{\mu\nu}\partial_\nu\phi$$

and the topological charge

$$P = \int i^0 dx$$

(here, $\epsilon^{\mu\nu}$ is a unit antisymmetric tensor).

Lagrangian (1) yields the following equations of motion:

$$\partial_\mu\partial^\mu\xi + h^2\phi^2\xi = 0, \tag{4}$$

$$\partial_\mu\partial^\mu\phi + 2h^2\xi\xi^*\phi + 2m^2(\phi^2 - v^2)\phi = 0. \tag{5}$$

The vacuum states of the system described by Lagrangian (1) are determined as follows:

$$\begin{aligned}\phi_{\text{vac}} &= \pm v,\\ \xi_{\text{vac}} &= 0.\end{aligned} \tag{6}$$

If we consider such configurations that the field ξ is identically equal to its vacuum value, then Eq. (4) holds identically, while Eq. (5) is rewritten as

$$\partial_\mu\partial^\mu\phi + 2m^2(\phi^2 - v^2)\phi = 0, \tag{7}$$

which coincides with the equation of motion for the Lagrangian of the $\lambda\phi^4$ theory with a doubly degenerate vacuum. In (1 + 1) dimensions, Eq. (7) has a static topological solution—a kink:

$$\phi_k(x) = v\tanh(mvx). \tag{8}$$

Since the field ξ in this theory is identically zero, the charge is also zero. The energy of this configuration (the mass of the kink) is given by

$$E_k = \frac{4}{3}mv^3. \tag{9}$$

Next, we will show that, for the field systems defined by Lagrangian (1) in the case of large charge Q, there exist localized configurations that are stable with respect to the decomposition into plane waves. These configurations may be either nontopological (a Q-ball) or topological (a topological Q-ball). Next, we will need explicit expressions for the energy of plane waves—perturbations of the vacuum states (6). In this case, to avoid diverging quantities while calculating the energy and charge, we consider these perturbations on a large (although finite) interval of the x axis: $-L/2 \le x \le L/2$. Under the assumption that the amplitudes of the field deviations from (6) are small, Eqs. (4) and (5) are rewritten in linear approximation as

$$\begin{aligned}\partial_\mu\partial^\mu\delta\xi + h^2v^2\delta\xi &= 0,\\ \partial_\mu\partial^\mu\delta\phi + 4m^2v^2\delta\phi &= 0,\end{aligned} \tag{10}$$

where $\delta\phi = \phi - \phi_{\text{vac}}$ and $\delta\xi = \xi - \xi_{\text{vac}}$. For a given charge of the configuration, a solution to system (10) that guarantees the minimal deviation of the energy from its vacuum value is given by

$$\begin{aligned}\delta\phi &= 0,\\ \delta\xi &= \sqrt{\frac{Q}{2hvL}}e^{ihvt}.\end{aligned} \tag{11}$$

The energy of solution (11) for the given charge Q is

$$E^0 = hvQ. \tag{12}$$

The essential difference of the case of one space dimension from the three-dimensional case is the fact that the system with Lagrangian (1) has topological solutions. When discussing the possibility of the existence of topological charged solutions, we should compare the energy of such a solution, which carries a charge Q, with the sum of the energy of the kink (8) (which has the least energy among all topological configurations) and the energy of the charged nontopological configuration (11). The value of this sum is

$$\tilde{E}^0 = hvQ + \frac{4}{3}mv^3. \tag{13}$$

It is this value that we take below as the starting point when discussing the existence of a stable charged topological configuration.

Now, following [7, 10], consider the energy functional

$$\begin{aligned}H[\phi,\xi] = \int dx\Bigg(&\left|\frac{\partial\xi}{\partial t}\right|^2 + \left|\frac{\partial\xi}{\partial x}\right|^2 + \frac{1}{2}\left(\frac{\partial\phi}{\partial t}\right)^2\\ &+ \frac{1}{2}\left(\frac{\partial\phi}{\partial x}\right)^2 + h^2\phi^2\xi\xi^* + \frac{m^2}{2}(\phi^2 - v^2)^2\Bigg)\end{aligned} \tag{14}$$

on the class of trial functions of the form

$$\phi(x) = \begin{cases}0, & |x| < l/2\\ \pm v\left[1 - \exp\left(\dfrac{x + l/2}{a}\right)\right], & x \le -l/2\\ v\left[1 - \exp\left(-\dfrac{x - l/2}{a}\right)\right], & x \ge l/2,\end{cases} \tag{15}$$

$$\xi(x,t) = \begin{cases}0, & |x| > l/2\\ A\cos(\Omega x)\exp(i\Omega t), & |x| \le l/2,\end{cases} \tag{16}$$

where l and a are arbitrary positive constants that should be varied (the Ritz parameters). The signs "+" and "–" in Eq. (15) correspond to nontopological and topological trial functions, respectively. We consider the state with the minimal energy such that $|\xi(x)|$ does not vanish on the interval $|x| < l/2$; therefore,

$$\Omega = \frac{\pi}{l}.$$

When $\xi(x, t) = |\xi(x)|\exp(i\Omega t)$, the charge of the configuration is given by

$$Q = 2\Omega\int|\xi(x)|^2 dx.$$

Hence, we can determine the constant A in Eq. (16):

$$A = \sqrt{\frac{Q}{\pi}}.$$

Now we can write out an explicit expression for the energy of the trial functions (15) and (16):

$$E = \frac{\pi Q}{l} + \frac{v^2}{2a} + \frac{m^2 v^4 l}{2} + \frac{11am^2 v^4}{12}. \qquad (17)$$

This expression does not contain the fields coupling constant h because this quantity enters in the Hamiltonian only in the combination $h^2\phi^2\xi\xi^*$, which vanishes identically for the functions of the form (15) and (16). Note also that expression (17) holds for nontopological as well as for topological configurations because the field ϕ enters in the Lagrangian quadratically and the change of sign in Eq. (15) does not affect the value of the energy.

The minimal value of expression (17) for a given charge Q is attained for the following values of the Ritz parameters:

$$l_0 = \frac{\sqrt{2\pi Q}}{mv^2}, \quad a_0 = \frac{1}{mv}\sqrt{\frac{6}{11}}$$

and is equal to

$$E_{\min} = mv^2\sqrt{2\pi Q} + mv^3\sqrt{\frac{11}{6}}. \qquad (18)$$

Since the energy of plane-wave configurations is a linear function of charge (see (12) and (13)), it turns out that the localized configuration (15), (16), whose energy depends on Q as $\sqrt{Q}$ is energetically favorable for large values of the charge. This means that, for large $Q > Q^{cr}$, there exists a localized solution in system (1). Note that the energy (13) of the topological configuration "kink + plane waves" with a given charge Q is different from energy (12) for a nontopological plane-wave configuration with the same charge Q. At the same time, the variational estimate (18) is equally applicable to nontopological and topological configurations. Hence, we obtain different expressions[1] for the critical charge Q^{cr}:

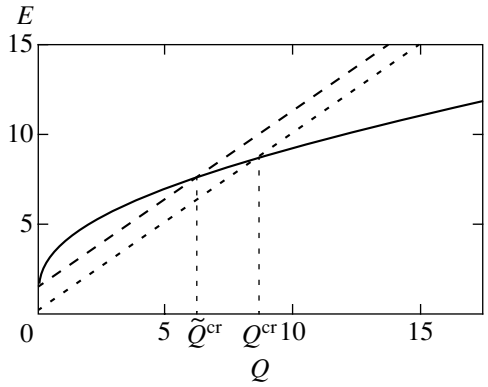

Fig. 1. The energy as a function of charge (dotted curve) for the plane-wave configuration, (dashed curve) for the "kink + plane waves"-type configuration, and (solid curve) for the variational configuration (15), (16). Q^{cr} and $\tilde{Q}^{cr}$ are critical values of charge for the nontopological and topological configurations, respectively. The graphs are drawn for the constants $h = m = 1$ and $v = 1$.

(1) nontopological Q-ball,

$$Q^{cr} = \frac{m^2 v^2}{h^2}\left(\pi + \frac{h}{m}\sqrt{\frac{11}{6}} + \sqrt{\pi^2 + 2\pi\frac{h}{m}\sqrt{\frac{11}{6}}}\right), \qquad (19)$$

(2) topological Q-ball,

$$\tilde{Q}^{cr} = \frac{m^2 v^2}{h^2}\left(\pi + \frac{h}{m}\left(\sqrt{\frac{11}{6}} - \frac{4}{3}\right) + \sqrt{\pi^2 + 2\pi\frac{h}{m}\left(\sqrt{\frac{11}{6}} - \frac{4}{3}\right)}\right). \qquad (20)$$

The meaning of the critical charges Q^{cr} and $\tilde{Q}^{cr}$ is clarified in Fig. 1. Thus, a topological configuration becomes energetically favorable and, hence, stable with respect to the decomposition into a kink + plane waves under critical charges less than those for a nontopological configuration.

3. AN EXACT SOLUTION FOR A TOPOLOGICAL Q-BALL

In the preceding section, we demonstrated that, in addition to an ordinary Q-ball, there can exist a topological Q-ball that carries topological and $U(1)$ charges simultaneously in a system defined by Lagrangian (1). However, the topological solution presented in the preceding section is variational. Actually, an exact solution to a problem with given boundary conditions at infinity

[1] These values represent upper estimates obtained with the use of trial functions (15) and (16); therefore, stable charged localized solutions may actually exist even for the charges less than (19) and (20).

(a given topological sector) and the total $U(1)$ charge Q may have even lower energy. It will be shown below that a particular solution to this problem for Eqs. (4) and (5) can be found analytically.

Indeed, we will seek a solution to Eqs. (4) and (5) in the form

$$\begin{aligned} \phi(x, t) &= \phi(x), \\ \xi(x, t) &= f(x)e^{i\omega t}, \end{aligned} \tag{21}$$

where $\phi(x)$ and $f(x)$ are real functions of x. Taking into account (21), we can rewrite Eqs. (4) and (5) as

$$-\omega^2 f - f'' + h^2\phi^2 f = 0, \tag{22}$$

$$-\phi'' + 2h^2 f^2\phi + 2m^2(\phi^2 - v^2)\phi = 0. \tag{23}$$

Next, we restrict the analysis to the assumption that the field ϕ preserves the functional dependence of the kink (8) (possibly, with a different scale):

$$\phi(x) = v\tanh(\alpha m v x), \tag{24}$$

where $\alpha > 0$ is a certain unknown constant.

Let us pass to a new variable

$$z = v\tanh(\alpha m v x), \quad -v < z < v.$$

in Eqs. (22) and (23). Then, we obtain

$$\begin{aligned} -\omega^2 f - \alpha^2 m^2 (v^2 - z^2)\frac{d}{dz}\left((v^2 - z^2)\frac{df}{dz}\right) \\ + h^2\phi^2 f = 0, \end{aligned} \tag{25}$$

and

$$\begin{aligned} -\alpha^2 m^2 (v^2 - z^2)\frac{d}{dz}\left((v^2 - z^2)\frac{d\phi}{dz}\right) \\ + 2h^2 f^2\phi + 2m^2(\phi^2 - v^2)\phi = 0 \end{aligned} \tag{26}$$

instead of (22) and (23). Taking into account that $\phi(z) = z$ in view of (24), from (26) we obtain

$$f^2(z) = \frac{m^2(1 - \alpha^2)}{h^2}(v^2 - z^2), \tag{27}$$

hence, the sought-for solution exists even if the constant α less than unity. Supposing that this condition is fulfilled, we find a root in (27) and substitute the expression for $f(z)$ thus obtained into (25). Then, we obtain

$$(2\alpha^2 m^2 - h^2)z^2 + (\omega^2 - \alpha^2 m^2 v^2) = 0. \tag{28}$$

Since equality (28) must hold for arbitrary z, we have

$$\alpha^2 = \frac{h^2}{2m^2}, \quad \omega^2 = \alpha^2 m^2 v^2 = \frac{h^2 v^2}{2}.$$

Thus, we have found an exact nontrivial solution to Eqs. (4) and (5). Its explicit expression is

$$\phi(x) = v\tanh(hvx/\sqrt{2}), \tag{29}$$

$$\xi(x, t) = v\sqrt{\frac{m^2}{h^2} - \frac{1}{2}}\frac{\exp(ihvt/\sqrt{2})}{\cosh(hvx/\sqrt{2})}. \tag{30}$$

This solution exists under the following constraints imposed on the parameters of the Lagrangian:

$$\frac{h^2}{m^2} \equiv \rho^2 \le 2. \tag{31}$$

Note that the spatial scale of the field ϕ in solution (29) and (30) coincides with the scale of the field ξ, and, as $\rho \longrightarrow \sqrt{2}$, solution (29) and (30) reduces to the uncharged topological kink (8). Let us write out the energy and charge of solution (29) and (30):

$$E = \frac{2\sqrt{2}}{3}hv^3\left(4\frac{m^2}{h^2} - 1\right), \tag{32}$$

$$Q = 2v^2\left(2\frac{m^2}{h^2} - 1\right). \tag{33}$$

Now, let us discuss how the solution obtained and the corresponding energy and charge are changed under the variation of the parameters of the Lagrangian. When $h \ll m$, the characteristic spatial scale of solution (29) and (30) is much greater than the size of the kink (8); in this case, the amplitude of the charged field is large ($\sim vm/h$). It follows from relations (32) and (33) that the energy and charge of a topological Q-ball are also large in the limit of a small coupling constant h. The tendency of the amplitude ξ in (30) to infinity as $h \longrightarrow 0$ implies that solution (29) and (30) is nonperturbative with respect to the small parameter h. This solution cannot be obtained by the expansion in powers of h for small h (when $h \longrightarrow 0$, the interaction term in (1) is of the order of m^2v^4, i.e., is not small as compared with other terms). On the contrary, the weak-coupling limit in (29) and (30) corresponds to the case $h \longrightarrow \sqrt{2}\,m$. In this limit, the amplitude of ξ in (30) as well as the charge tend to zero, whereas $\phi(x)$ tends to $\phi_k(x)$ (8). Thus, the weak-coupling limit corresponds to the smallness of $h^2|\xi|^2$ rather than of h^2. Here, we note that the exact topological solution essentially depends on the constant h in contrast to the variational solution proposed above, which did not depend on h.

Now, let us compare the energy of configuration (29), (30) with the energies of a plane-wave configuration and the corresponding trial configuration (15) and

(16). For this purpose, we first should substitute expression (33) for the charge Q into (13) and (18). This yields

$$\widetilde{E}^0 = mv^3\left[\frac{4}{3} + 2\rho\left(\frac{2}{\rho^2} - 1\right)\right], \quad (34)$$

$$E_{\min} = mv^3\left[2\sqrt{\pi\left(\frac{2}{\rho^2} - 1\right)} + \sqrt{\frac{11}{6}}\right]. \quad (35)$$

One can easily verify that both the energy (34) of the plane-wave configuration and the energy (35) of the variational configuration with an appropriate charge prove to be greater than the energy (32) of the exact solution for any values of the parameters of the Lagrangian that satisfy inequality (31). This fact is natural just as the fact that the charge Q (33) is less than the estimated critical value (20) for the topological configuration since (29), (30) is an exact solution to the equations of motion. Thus, the exact solution (29), (30) is stable with respect to the decomposition into a kink + plane waves and into configurations close to the variational ones (15), (16).

Figure 2 represents the ratio of energy to the constant m in the Lagrangian versus ρ in the interval $0 < \rho \le \sqrt{2}$ for the exact solution, for the "kink + plane waves"-type configuration, and for the variational configuration (15), (16).

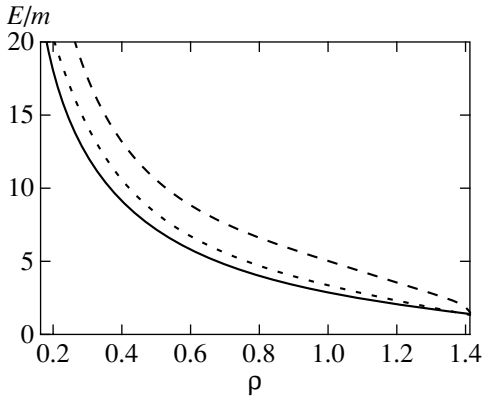

Fig. 2. The ratio of energy to the constant m versus ρ (solid curve) for the exact solution, (dotted curve) for the "kink + plane waves"-type configuration, and (dashed curve) for the variational configuration (15), (16). In the last two cases, the charge Q is equal to the charge (33) of the exact solution. The graphs are drawn for the constant $v = 1$.

4. IMPROVED VARIATIONAL PROCEDURE

The exact solution (29), (30) with the energy less than that of the solution obtained with the use of the variational procedure described in Section 2 indicates that this variational procedure can and must be improved. Moreover, for given parameters of the Lagrangian, the charge (33) of the exact solution is fixed; therefore, the exact solution does not reproduce the entire spectrum of topological solutions with different charges Q for given parameters of Lagrangian (1). Let us try to improve the variational procedure. For this purpose, we take the variational functions in the following form:

$$\phi(x) = v\tanh(\beta x), \quad (36)$$

$$\xi(x, t) = Av\frac{\exp(i\beta t)}{\cosh(\beta x)}, \quad (37)$$

where A and β are the Ritz parameters. This choice of the variational function with

$$A = \sqrt{\frac{m^2}{h^2} - \frac{1}{2}}, \quad \beta = \frac{hv}{\sqrt{2}}$$

reproduces the exact solution (29), (30). At the same time, one can hope that, since the variational functions continuously depend on the parameters A and β the substitution of (36) and (37) into the energy functional (14) will result in better minimization of this functional at least for the values of A and β close to those for the exact solution. The substitution of (36) and (37) into (14) yields the following expressions for the energy and charge of this configuration:

$$\tilde{E} = \frac{2}{3}v^2(1 + 4A^2)\beta + \frac{2}{3}v^4(h^2A^2 + m^2)\frac{1}{\beta}, \quad (38)$$

$$Q = 4A^2v^2. \quad (39)$$

Thus, the value of A is uniquely specified by the charge. Substituting the value of A obtained from (39) into (38) and minimizing the energy with respect to β for a given charge, we obtain the following expression for the minimal value of energy (38):

$$\tilde{E}_{\min} = \frac{2}{3}v\sqrt{(Q + v^2)(h^2Q + 4m^2v^2)}, \quad (40)$$

this minimum is attained for the value of β given by

$$\beta_0 = \frac{v}{2}\sqrt{\frac{h^2Q + 4m^2v^2}{Q + v^2}}.$$

When Q is equal to the charge (33) of the exact solution, we obtain, as expected,

$$\beta_0 = \frac{hv}{\sqrt{2}}, \quad \tilde{E}_{\min} = E = \frac{2\sqrt{2}}{3}hv^3\left(4\frac{m^2}{h^2} - 1\right).$$

As we have already mentioned, when the charges differ little from Q (33), the energy (40) of the new variational function is less than the energy (18) of configuration (15), (16) as well as than the energy (13) of the "kink + plane waves"-type configuration. The behavior

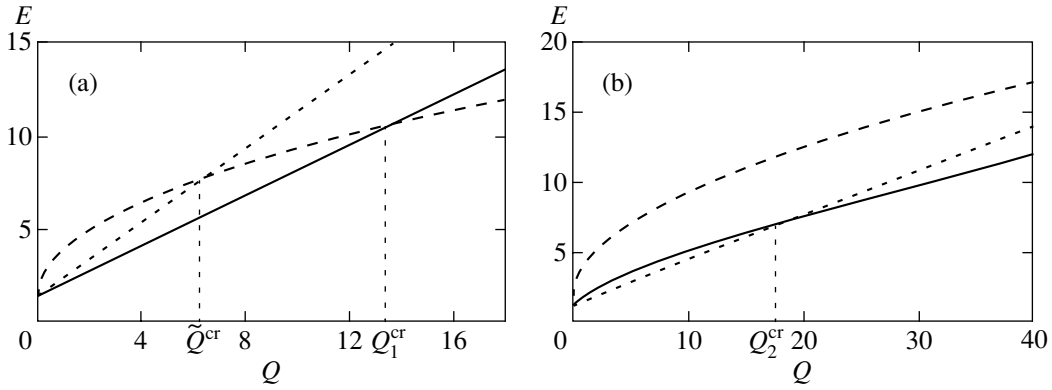

Fig. 3. The energy as a function of charge for the variational configurations (dashed curve) (15), (16) and (solid curve) (36), (37) and (dotted curve) for the "kink + plane waves"-type configuration that correspond to (a) $\rho = 1$ ($h = m = 1$) and $v = 1$ and (b) $\rho = \sqrt{0.1} = 0.31622\ldots$ ($h = \sqrt{0.1}$, $m = 1$) and $v = 1$.

of expression (40) for arbitrary values of the charge is as follows.

(1) When $Q \longrightarrow +\infty$,

$$\tilde{E}_{\min} = \frac{2}{3} h v Q + O(\sqrt{Q});$$

hence, when the charge is greater than a certain Q_1^{cr} (this value of Q_1^{cr} can be obtained in the general case from the equality between $\tilde{E}_{\min}(Q_1^{cr})$ (40) and $E_{\min}(Q_1^{cr})$ (18)), the variational configuration (15), (16) becomes energetically more favorable as compared with (36), (37); thus, for the large charge of a system, the conventional substitution that leads to a *Q*-ball-type solution proves to be energetically more favorable.

(2) When $Q < Q_1^{cr}$, there are two possible cases depending on the ratio $\rho = h/m$ of the Lagrangian constants.

(a) The stability condition of a configuration with respect to the decomposition into a kink + plane waves is satisfied for any $Q > 0$. This situation occurs when $3 - \sqrt{5} \le \rho \le 3 + \sqrt{5}$. The characteristic form of the function of energy versus charge for both the variational and the plane-wave configurations is shown in Fig. 3a.

(b) The equality between the energies of the variational and plane-wave configurations, $\tilde{E}_{\min}(Q) = \tilde{E}^0(Q)$, holds for $Q > 0$, which corresponds to $\rho > 3 + \sqrt{5}$ or $0 < \rho < 3 - \sqrt{5}$. In this case, the configuration (35), (36) with charges *Q* smaller than

$$Q_2^{cr} = \frac{4}{5} v^2 \left(\frac{4}{\rho^2} - \frac{6}{\rho} + 1 \right)$$

is energetically less favorable as compared to the plane-wave configuration; therefore, we cannot draw any conclusions about the existence of stable localized solutions. The characteristic layout of energy-versus-charge curves for this case is shown in Fig. 3b.

However, it is essential that, according to the conclusion made above, the energy (32) of the exact solution is less than the energy (34) of the "kink + plane waves"-type configuration with the corresponding charge (33) for any $0 < \rho \le \sqrt{2}$. Thus, the variational functions (36), (37) give an exact solution for the following values of the charge: $Q = 0$ and $Q = 2v^2(2/\rho^2 - 1)$. In this case, for any $Q > Q_2^{cr}$ (or $Q \ge 0$, depending on the value of ρ), the functions (36) and (37) yield a lesser value of the energy functional than the "kink + plane waves"-type configuration, whereas, for $Q < Q_1^{cr}$, they yield a lesser value of the energy functional than the variational configuration (15), (16).

Thus, we have determined the variational function for the values of the charge lesser than the critical value $\tilde{Q}_1^{cr}$ for the functions (15) and (16); the new variational functions coincide with the exact solution (29), (30) for an appropriate value of the charge. The exact solution (29), (30) and the improved variational procedure (36), (37) imply that topological *Q*-balls can exist in a much wider range of charges *Q* than are implied by the standard variational procedure usually applied to the *Q*-balls.

5. CONCLUSION

We have demonstrated that a wall-type one-dimensional solution (a kink) for a scalar Higgs field can bind a complex scalar field carrying a $U(1)$ charge and that the value of the charge *Q* bound by the wall may vary within rather wide limits.

The possibility of binding a large $U(1)$ charge by a topological wall is a new fundamental result obtained in this paper. This phenomenon should manifest itself in

the problem of particle scattering by a domain wall. Such a problem was first discussed in Voloshin's paper [2]. He considered two different problems. The first problem dealt with the scattering of the Higgs plane wave by the kink of this very field. In this case, the scattering potential proved to be reflectionless. The second problem dealt with the interaction between the Higgs field and a fermion field. The problem of scattering of a fermion by a kink was also reduced to the problem of scattering in a given potential induced by the field of this kink. The statement of this problem within the framework of [2] corresponds in our case to the study of the scattering problem for Eq. (4) with a given potential $\phi(x)$. The substitution of $\phi(x)$ in form (8) reduces the problem to the known problem of the scattering of a particle in the field $V(x) \propto -1/\cosh^2(mvx)$. Interestingly, as $h \longrightarrow 0$, the reflection coefficient of such a potential tends to unity as $k \longrightarrow 0$ [11]. Such behavior of the reflection coefficient for $h \longrightarrow 0$ implies that the weak-coupling approximation is inapplicable. This result is not incidental in the problem considered because there always exists a bound state in the case of attraction potential in the one-dimensional case.

Note that, generally, the problem of the scattering of particles by a kink is not reduced to the problem of scattering by a given potential. In fact, in addition to Eq. (4), which describes the behavior of the field ξ in the given field $\phi(x)$, one has to analyze the behavior of the field $\phi(x)$ in the field $\xi(x)$ of a plane wave, i.e., analyze the solution to Eq. (5). As is clear from (5), when one takes $\xi(x)$ in the form of a plane wave, the vacuum of the field ϕ and the excitation mass of the field ϕ over the vacuum are changed. Thus, the problem of scattering of particles in the field of a kink is self-consistent only if the particles represent small deviations of this field from the vacuum. Moreover, even the bound state of the field ξ in the potential of the kink of the field ϕ can be obtained without taking into account the reaction of the field ξ on the field ϕ formally only in the case of a sufficiently strongly coupled level. Qualitatively, this result becomes clear if we take into account that the wave function of a weakly coupled level slowly decays at large distances; therefore, the term $2h^2\xi\xi^*\phi$, which is rejected in (5) for $\phi(x) = v\tanh(mvx)$, becomes exponentially large as compared with the term $(\phi^2 - v^2)\phi \propto \cosh^{-2}(mvx)$ as $x \longrightarrow \pm\infty$. When there is a single bound level in the potential $-h^2v^2\cosh^2(mvx)$, the wave function of this level is given by [11]

$$\xi = \frac{e^{i\omega t}}{\cosh^s(mvx)},$$

where

$$s = \frac{1}{2}(-1 + \sqrt{1 + 4\rho^2}).$$

To prevent the term $2h^2\xi\xi^*\phi$ from being exponentially large for large x as compared with other terms in (5), one should satisfy the inequality $s \geq 1$; hence, we obtain $\rho \geq \sqrt{2}$. Thus, the reaction of the field ξ on the field ϕ should be taken into account not only when solving a scattering problem but also when solving the problem of a weakly coupled particle of the field ξ in the field of a kink.

The existence of topological walls carrying a $U(1)$ charge may be directly related to the problem of the collapse of domain bubbles. The problem of evolution of a collapsing domain bubble has been studied in detail and from various viewpoints starting from the pioneering work by Zel'dovich *et al.* [12] (see also [13] in this relation). Usually, one considers the evolution of spherically symmetric bubbles of the $\lambda\phi^4$ theory whose field profile along the radius corresponds to the profile of a kink. As is known, such bubbles would collapse; therefore, there do not exist stationary domain structures in the $\lambda\phi^4$ theory. The possibility of stabilizing domain bubbles by filling them with charges has been discussed starting from [14, 15]. In particular, Bardeen *et al.* [15] discussed the stabilization of bubbles by filling them with quarks. The solutions for charged walls obtained in the present paper can also be used for finding stable domain regions carrying a $U(1)$ charge. Here, we should stress that a vacuum domain surrounded by a charged wall may, in principle, evolve into a Q-ball. In future, we are planning to revert to the question of the time evolution of such configurations.

The topological solutions with a $U(1)$ charge obtained in this study may possibly exist even in certain more general models. In particular, a hedgehog-type solution for a triplet of Higgs scalar fields is of interest when one takes into account its interaction with a complex field with $U(1)$ symmetry.

ACKNOWLEDGMENTS

We are grateful to N.A. Voronov, N.B. Konyukhova, and S.V. Kurochkin for useful discussions.

This work was supported in part by the Russian Foundation for Basic Research, project no. 00-15-96562.

REFERENCES

1. A. Kovner, M. Shifman, and A. Smilga, Phys. Rev. D **56**, 7978 (1997); G. Dvali and M. Shifman, Phys. Lett. B **396**, 64 (1997); A. Hanany and K. Hori, Nucl. Phys. B **513**, 119 (1998); E. Witten, Nucl. Phys. B **507**, 658 (1997); J. Gauntlett, R. Portugues, D. Tong, and P. Townsend, Phys. Rev. D **63**, 085002 (2001).
2. M. B. Voloshin, Yad. Fiz. **21**, 1331 (1975) [Sov. J. Nucl. Phys. **21**, 687 (1975)].
3. G. R. Farrar and J. W. McIntosh, Phys. Rev. D **51**, 5889 (1995).

4. G. Dvali, H. Liu, and T. Vachaspati, Phys. Rev. Lett. **80**, 2281 (1998).
5. A. E. Kudryavtsev, B. M. A. G. Piette, and W. J. Zakrzewski, Phys. Rev. D **61**, 025016 (2000).
6. J. F. Gomes, E. P. Gueuvoghlanian, G. M. Sotkov, and A. H. Zimerman, hep-th/0007169.
7. R. Friedberg, T. D. Lee, and A. Sirlin, Phys. Rev. D **13**, 2739 (1976).
8. R. A. Leese, Nucl. Phys. B **366**, 283 (1991).
9. S. Coleman, Nucl. Phys. B **262**, 263 (1985).
10. V. A. Rubakov, *Classical Gauge Fields* (Editorial URSS, Moscow, 1999).
11. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory* (Nauka, Moscow, 1989, 4th ed.; Pergamon, New York, 1977, 3rd ed.s).
12. Ya. B. Zel'dovich, I. Yu. Kobzarev, and L. B. Okun', Zh. Éksp. Teor. Fiz. **67**, 3 (1974) [Sov. Phys. JETP **40**, 1 (1975)].
13. V. A. Kuzmin, V. A. Rubakov, and M. E. Shaposhnikov, Phys. Lett. B **155B**, 36 (1985); G. Farrar and M. E. Shaposhnikov, Phys. Rev. Lett. **70**, 2833 (1993).
14. P. Viciarelly, Nuovo Cimento Lett. **4** (16), 905 (1972).
15. W. A. Bardeen, M. S. Chanowitz, S. D. Drell, *et al.*, Phys. Rev. D **11**, 1094 (1975).

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 93, No. 4, 2001, pp. 685–694.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 786–796.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Photogeneration of Neutrinos and Axions under the Stimulating Effect of a Strong Magnetic Field

V. V. Skobelev

Moscow State Industrial University, Moscow, 109280 Russia
e-mail: skobelev@mail.msiv.ru

Received May 3, 2001

Abstract—Photogeneration of neutrinos and axions at nuclei, $\gamma(Ze) \longrightarrow \gamma(\nu\bar{\nu})$, γa, and inelastic photon–photon scattering, $\gamma\gamma \longrightarrow \gamma(\nu\bar{\nu})$, γa, are considered in the 2D covariant theory being developed for calculating matrix elements of Feynman diagrams in a strong magnetic field. Since the matrix elements of four-pole diagrams are linear functions of the magnetic induction B, the contribution of the radiative photogeneration of neutrinos at nuclei to the luminosity of magnetic neutron stars at early stages of their evolution may compete with URCA processes for values of $B \sim (10^3–10^4)B_0$ ($B_0 = m_e^2/|e| = 4.41 \times 10^{13}$ G). The upper estimate of the axion mass obtained from the condition of dominance of the neutrino luminosity over the axion luminosity for the proposed values of temperature and magnetic induction is in accord with other independent results. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is generally accepted [1, 2] that in an extreme astrophysical situation of the type of a supernova burst accompanied with the formation of a neutron star, neutrino emission is the dominating mechanism of energy release in view of the high penetrability of neutrinos. Naturally, neutrino physics itself is an important component of the theory of elementary particles. Important aspects of neutrino physics are the problem of the existence of neutrinos with definite masses [3, 4], the form of the unitary matrix of mixing during the formation of neutrino states ν_e, ν_μ, and ν_τ constituting weak current, the number of neutrino flavors, etc. In this respect, collapsed astrophysical objects as well as the Universe as a whole with the possible dominance of the massive neutrino component as a carrier of the latent mass are giant natural laboratories enhancing our understanding of the nature of neutrino.

Apart from standard nuclear reactions (URCA processes), quantum processes of the Compton type $\gamma e^- \longrightarrow e^-(\nu\bar{\nu})$ [5] and bremsstrahlung $e^-(Ze) \longrightarrow e^-(\nu\bar{\nu})$ [6] also contribute to the formation of neutrino radiation by the above-mentioned processes. The possible generation of neutrinos accompanied by the excitation of vacuum electron states was apparently mentioned by Rosenberg [7] who studied the conversion of a photon to a pair of neutrinos at a nuclei, $\gamma(Ze) \longrightarrow (\nu\bar{\nu})$, in the Fermi diagram, which is determined by the contribution of the electron tripole. A noticeable contribution from such a mechanism is possible at a high temperature $T \sim m$ (m is the electron mass) of equilibrium photon radiation and a high concentration of nuclei. This appears astonishing in itself since such electrodynamic effects of a higher order with an electron quadripole were identified with difficulty in the laboratory conditions (the Delbruck effect [8, 9] and splitting of a photon at a nucleus [10, 11]), while the main nonlinear electrodynamic effect of light scattering by light has not yet been detected.

The existence of ultrastrong magnetic fields formed upon the compression of the primary magnetosphere with the magnetic flux conservation revealed basically new features in the formation of neutrino radiation during the collapse. According to estimates, the magnetic induction may attain the characteristic Schwinger value

$$B_0 = \frac{m^2}{|e|} = 4.41 \times 10^{13} \text{ G}$$

($e < 0$ is the electron charge) and even exceed it up to $B \sim 10^4 B_0$ [12]. This leads to the opening of other channels of neutrino generation, e.g., the $\gamma \longrightarrow (\nu\bar{\nu})$ channel [13, 14] or the synchrotron mechanism of neutrino emission, $e^- \longrightarrow e^-(\nu\bar{\nu})$ [15–18]. It should be noted, in particular, that a new possibility of neutrino mass detection on the basis of analysis of threshold effects in the synchrotron emission of massive neutrinos was indicated in [18]. In the following analysis, we will proceed from the average limitation on the electron neutrino mass, given in [19]:

$$m_{\nu_e} < 15 \text{ eV}. \quad (1)$$

A more stringent limitation does not affect the results obtained by us disregarding the neutrino mass as com-

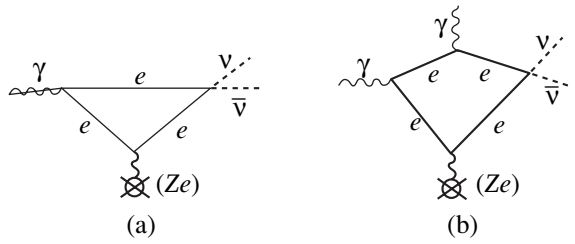

Fig. 1.

pared to the characteristic energies appearing in this work.

The effect of the field on the channels which are open in zero magnetic field also is the strongest when the real and virtual electrons are on the ground Landau level. This results in a transition to the 2D case of the theory in the (0, 3) subspace (the axis 3 is directed along the field). The corresponding mathematical apparatus was proposed in [14, 20] and developed in other publications which are mentioned in the literature cited here. Since this circumstance is ignored systematically by some authors (see, for example, [21]), we will briefly describe in Section 2 the basic principles of this approach and some of its main consequences under certain limitations imposed on the momenta of the external lines.

In particular, the matrix elements of diagrams without excitation of the vacuum (e.g., the diagram of the Compton effect) do not contain any field dependence and the role of the field is reduced only to the dependence of the "motion" of an electron on time and on one spatial coordinate. In this case, however, the field dependence leading to the stimulated enhancement of effect in some cases may appear in the integration with respect to quasimomenta of the outer electron lines as well as in the inclusion of the temperature distribution functions for the electron gas with the Fermi momentum defined as [22]

$$p_F = \frac{2\pi^2 n}{\gamma}, \tag{2}$$

where $\gamma = |eB|$ and n is the electron concentration. This is demonstrated, for example, in [23], where the Compton mechanism of emission of neutrinos and axions, $\gamma e^- \longrightarrow e^-(\nu\bar{\nu})$, $e^- a$ in a 2D magnetized Fermi gas is considered.

The stimulating effect of an external magnetic field with induction $B \gg B_0$ is also manifested in diagrams without outer electron lines. The matrix element is a linear function of the field for an even number of vectorial and (or) pseudovectorial vertices in an electron loop and becomes constant for an odd number of vertices. Consequently, the cross sections, probabilities, and emission power corresponding to the first type of processes contain the factor $(B/B_0)^2 \gg 1$ in contrast to the characteristics of the second type of processes and may exceed their values although these quantities have a higher order of magnitude in the formal perturbation theory. This statement does not contain any contradiction since, like in ordinary quantum electrodynamics (QED), the contribution of the nth order loop diagram with an even number of vectorial vertices exceeds the contribution of the $(n-1)$th diagram which is equal to zero (Farry theorem). The characteristics of the process of photogeneration of neutrino at a nucleus, $\gamma(Ze) \longrightarrow (\nu\bar{\nu})$, taking into account the contribution of the triangular diagram in a strong magnetic field $B \gg B_0$ (Fig. 1a) were analyzed in [24] using the Fermi scheme. It can be seen that it is expedient to analyze the radiation-induced generation of a neutrino pair, $\gamma \longrightarrow \gamma(\nu\bar{\nu})$, stimulated by the electron quadripole (Fig. 1b). It should be noted that similar channels of inelastic scattering of "light by light," $\gamma\gamma \longrightarrow \gamma(\nu\bar{\nu})$, and the fusion of photon at a nucleus, $\gamma\gamma(Ze) \longrightarrow (\nu\bar{\nu})$, as well as their contribution to luminosity, were studied by us earlier [25, 26].

The idea concerning the existence of a pseudoscalar Goldstone boson (axion) has been discussed recently. It makes it possible to explain the observed exact CP invariance of strong interactions [27]. An axion also competes with a neutrino as a potential carrier of latent mass and as one of the mechanisms of the energy expulsion from the star during its collapse in view of the smallness of the coupling constant $1/f$ for an axion with "ordinary" particles (invisible axion). For this reason, the program of studying simultaneously the neutrino and axion luminosities of stars in view of different mechanisms of their generation seems to be quite natural. Since the axion mass is strongly associated with the energy scale f of violation of the Peccei–Quinn global symmetry [27],

$$m_a \approx 0.6 \times 10^{-3}\ \text{eV}\left(\frac{10^{10}\ \text{GeV}}{f}\right), \tag{3}$$

a comparison of the neutrino and axion luminosities makes it possible to refine the possible range of axion mass, which is still quite wide:

$$10^{-5}\ \text{eV} \lesssim m_a \lesssim 10^{-2}\ \text{eV}. \tag{4}$$

In this connection, we will consider a channel of radiation-induced photogeneration of an axion at a nucleus, $\gamma(Ze) \longrightarrow \gamma a$ (Fig. 2), and an inelastic scattering channel $\gamma\gamma \longrightarrow \gamma a$ (Fig. 2b), which are also described by a quadripole diagram.

The paper is constructed as follows. In Section 2, the basic principles of the method of analysis of Feynman diagrams in strong magnetic fields, which was developed by us, are formulated. The matrix elements of the radiation-induced photogeneration of an axion at a nucleus, $\gamma(Ze) \longrightarrow \gamma a$, and of a pair of neutrinos at a nucleus, $\gamma(Ze) \longrightarrow \gamma(\nu\bar{\nu})$, are determined and the corresponding emission powers are calculated under the

assumption of an equilibrium distribution of photons. In Section 4, a similar analysis is carried out for the inelastic scattering $\gamma\gamma \longrightarrow \gamma a$ and the result obtained earlier for the process $\gamma\gamma \longrightarrow \gamma(\nu\bar{\nu})$ is presented. The characteristics of these effects are analyzed in Section 5 taking into account their contributions to the axion and neutrino luminosities of collapsed objects and possible limitations imposed on the axion mass.

2. FUNDAMENTALS OF CALCULATION OF THE MATRIX ELEMENTS OF FEYNMAN DIAGRAMS IN A STRONG MAGNETIC FIELD

The solution of the Dirac equation in the Cartesian coordinates in a constant and uniform magnetic field,

$$(i\hat{\partial} - e\hat{A} - m)\Psi = 0, \tag{5}$$

$$A_\alpha = Bx_1 g_{\alpha 2}, \tag{5a}$$

was found in [29]. In the standard representation of γ matrices, it has the form

$$\Psi = \frac{(\gamma/\pi)^{1/4}}{(2p_0L_2L_3)^{1/2}} \times \exp\left[-\frac{\xi^2}{2} + i(p_2x_2 + p_3x_3)\right]u_n, \tag{6}$$

$$u_n = \frac{2^{-n/2}}{\sqrt{n!}}\begin{pmatrix} -iC_1\sqrt{2n}H_{n-1}(\xi) \\ C_2H_n(\xi) \\ -iC_3\sqrt{2n}H_{n-1}(\xi) \\ C_4H_n(\xi) \end{pmatrix}, \tag{7}$$

$$\xi = x_1\sqrt{\gamma} + \frac{p_2}{\sqrt{\gamma}}, \tag{8}$$

where $H_n(\xi)$ are the Hermite polynomials, and the spin coefficients C_j satisfy the normalization condition

$$\sum_{j=1}^{4} |C_j|^2 = 2p_0. \tag{9}$$

Here,

$$p_0 = \sqrt{m^2 + p_3^2 + 2\gamma n}$$

is the energy of an electron with momentum p_3 along the field and with the value of the quantum number $n = 0, 1, 2, \ldots$; p_2 is the quasimomentum characterizing the position of the packet center on the axis 1; and $L_{2,3}$ are the auxiliary normalization lengths along the axes 2 and 3.

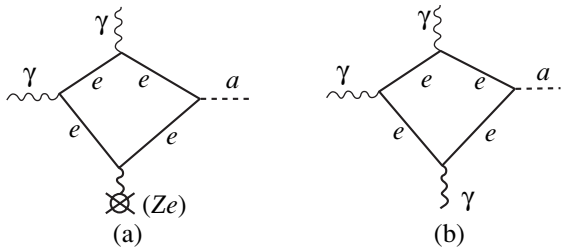

Fig. 2.

At the Landau ground level, $n = 0$, the electron spin is directed against the field and the spin coefficients are given by

$$C_1 = C_3 = 0, \quad C_2 = \frac{p_0 + m}{\sqrt{p_0 + m}}, \quad C_4 = -\frac{p_3}{\sqrt{p_0 + m}}, \quad p_0 = \sqrt{p_3^2 + m^2}. \tag{10}$$

The resultant spinor $u_0(p)$ is independent of the coordinate and satisfies the following equations:

$$\begin{cases} (\hat{p}_\parallel - m)u_0 = 0, \quad \hat{p}_\parallel = p_0\gamma^0 + p_3\gamma^3, \\ \Pi_- u_0 = u_0, \quad \Pi_- = \dfrac{1 - i\gamma_1\gamma_2}{2}, \\ \bar{u}_0u_0 = 2m, \quad u_0\bar{u}_0 = \Pi_-(\hat{p}_\parallel + m). \end{cases} \tag{11}$$

This means that at the Landau ground level, the space of physical momenta and γ matrices in fact degenerates into the 2D space (0, 3) so that we can introduce the 2D spinor $v(p)$ and the 2×2 matrices $\tilde{\gamma}_\alpha$ ($\alpha = 0, 3$) with the properties

$$(\check{p} - m)v = 0, \quad \check{p} = p_0\tilde{\gamma}^0 + p_3\tilde{\gamma}^3, \tag{12a}$$

$$\bar{v}v = 2m, \quad \bar{v} = v^+\gamma^0, \tag{12b}$$

$$v\bar{v} = (\check{p} + m). \tag{12c}$$

In this case,

$$\tilde{\gamma}^0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \tilde{\gamma}^3 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

in the representation equivalent to the standard representation, while expressions (12) are independent of representation.

The main reduction formula rendering the procedure of calculation of traces and convolutions trivial has the form

$$\tilde{\gamma}_\alpha\tilde{\gamma}_\beta = \tilde{g}_{\alpha\beta} + \tilde{\gamma}^5\varepsilon_{\alpha\beta}, \tag{13}$$

where $\tilde{\gamma}^5 = \tilde{\gamma}^0\tilde{\gamma}^3$, $(\tilde{\gamma}^5)^2 = 1$, $\tilde{g}_{\alpha\beta} = (1, -1)$ is the metric tensor and $\varepsilon_{\alpha\beta}$ ($\varepsilon_{30} = -\varepsilon_{03} = 1$) is the absolutely antisymmetric tensor in the subspace (0, 3). In particular, taking into account relation

$$\varepsilon^{\alpha\beta}\varepsilon^{\rho\sigma} = \tilde{g}^{\alpha\sigma}\tilde{g}^{\beta\rho} - \tilde{g}^{\alpha\rho}\tilde{g}^{\beta\sigma} \tag{14}$$

we can easily derive the following relations:

$$\tilde{\gamma}_\alpha\tilde{\gamma}_\beta + \tilde{\gamma}_\beta\tilde{\gamma}_\alpha = 2\tilde{g}_{\alpha\beta}, \tag{15a}$$

$$\tilde{\gamma}^\alpha(\tilde{\gamma}_{\alpha_1}\ldots\tilde{\gamma}_{\alpha_{2n+1}})\tilde{\gamma}_\alpha = 0, \tag{15b}$$

$$\frac{1}{2}\mathrm{Tr}(\tilde{\gamma}^5\tilde{\gamma}^\alpha\tilde{\gamma}^\beta\tilde{\gamma}^\rho\tilde{\gamma}^\sigma) = \tilde{g}^{\alpha\beta}\varepsilon^{\rho\sigma} + \varepsilon^{\alpha\beta}\tilde{g}^{\rho\sigma}, \tag{15c}$$

and so on.

The solution of the singular Dirac equation in a constant and uniform magnetic field, i.e.,

$$(i\hat{\partial} - e\hat{A} - m)S(x, y) = \delta(x - y), \tag{16}$$

in the same calibration (5a) can be presented in the form [24]

$$S(x, y) = f(x_\perp, y_\perp)G(x - y), \tag{17}$$

$$f(x_\perp, y_\perp) = \exp\left[-\frac{i\gamma}{2}(x_1 + y_1)(x_2 - y_2)\right], \tag{17a}$$

$$G(z) = \frac{1}{(2\pi)^4}\int d^4q e^{-i(qz)}G(q), \tag{17b}$$

$$G(q) = \frac{1}{\gamma\eta}\int_0^1 dt\left(\frac{1+t}{1-t}\right)^\eta \times e^{-\delta t}\left\{(\hat{q}_\parallel + m)\left[\Pi_-(1 - \delta t) - \frac{\eta}{1+t}\right] - \eta\hat{q}_\perp\right\}, \tag{17c}$$

where the following notation has been introduced:

$$\eta = \frac{q_\parallel^2 - m^2}{2\gamma}, \quad q_\parallel^2 = q_0^2 - q_3^2, \quad \delta = \frac{q_\perp^2}{\gamma}, \quad q_\perp^2 = q_1^2 + q_2^2, \quad \hat{q}_\perp = q_1\gamma^1 + q_2\gamma^2. \tag{18}$$

If the condition

$$|\eta| \ll 1 \tag{19}$$

is satisfied, Eq. (17c) assumes the form

$$G(q) = 2e^{-\delta}\Pi_-\frac{\hat{q}_\parallel + m}{q_\parallel^2 - m^2}.$$

Obviously, condition (19) is satisfied for diagrams without the excitation of vacuum if the momenta of external electron lines satisfy the condition

$$(p_i)_\parallel^2 - m^2 < 2\gamma, \tag{19a}$$

while the momenta of photon (axion, neutrino) lines satisfy the condition

$$(k_i p_j)_\parallel, (k_i)_\parallel^2 \ll \gamma. \tag{19b}$$

In the case of diagrams with excitation of the vacuum, the following additional constraint must be imposed in the case of convergence of integrals in the 2D (0, 3) momentum of the loop at the electron mass:

$$\gamma \gg m^2 \quad (B \gg B_0). \tag{19c}$$

Passing to 2D $\tilde{\gamma}$ matrices and convolutions in (0, 3), we present Green's function in the form

$$S_s(x, y) = \frac{\gamma}{2\pi}f(x_\perp, y_\perp)\varphi((x - y)_\perp) \times \frac{1}{(2\pi)^2}\int d^2q e^{-iq(x-y)}G_s(q), \tag{20}$$

$$\varphi(z_\perp) = \exp\left[-\frac{\gamma}{4}(z_1^2 + z_2^2)\right], \tag{20a}$$

$$G_s(q) = \frac{\hat{q} + m}{q^2 - m^2}, \tag{20b}$$

implying that the vertex factors in the matrix elements of the diagrams can also be transformed to 2D factors since

$$\Pi_-(\gamma_\alpha, \gamma_\alpha\gamma^5)\Pi_- \longrightarrow (\tilde{\gamma}_\alpha, \tilde{\gamma}_\alpha\tilde{\gamma}^5)$$

and differ from zero only for $\alpha = 0, 3$.

In the evaluation of the matrix elements of the diagrams without vacuum excitation (e.g., a diagram of the type of the Compton effect), the integral with respect to transverse coordinates has the form

$$J = \int dx_1 dx_2\int dy_1 dy_2 f(x_\perp, y_\perp)\varphi((x - y)_\perp) \times \exp\left[-\frac{1}{2}\left(\sqrt{\gamma}x_1 + \frac{p_2'}{\sqrt{\gamma}}\right)^2 - ip_2'x_2 - \frac{1}{2}\left(\sqrt{\gamma}y_1 + \frac{p_2}{\sqrt{\gamma}}\right)^2 + ip_2y_2 + ik_1y_1 + ik_2y_2 - ik_1'x_1 - ik_2'x_2\right].$$

After the evaluation of Gaussian type integrals, it has the form

$$J = 4\pi\left(\frac{\pi}{\gamma}\right)^{3/2}\delta(p_2 + k_2 - p_2' - k_2')\exp\left[-\frac{k_\perp^2 + k_\perp'^2}{4\gamma} + \frac{i}{2\gamma}(k_2k_1' - k_1k_2') + \frac{i}{2\gamma}(p_2 + p_2')(k_1' - k_1)\right], \tag{21}$$

where p_2 and p_2' are quasimomenta of the entering and emerging electron lines, and $k_{1,2}$ and $k_{1,2}'$ are the transverse momenta of the entering and the emerging lines of zero-charge particles (like photons). In view of con-

straints (19a) and (19b), the first and second terms in the exponent in Eq. (21) can be omitted, while the third term has the same form for other diagrams of the process also and can also be omitted as an insignificant phase factor in the case of real external zero-charge line. For example, the matrix element of a "2D" Compton effect taking into account formulas (6) and (20) can be written in the form

$$\langle f|S|i\rangle = i(2\pi)^3\delta^{(0,2,3)}(p+k-p'-k') \times \frac{M}{(2k_0 2k_0' 2p_0 2p_0')^{1/2} L_2 L_3 V}, \tag{22}$$

$$M = 4\pi\alpha\bar{v}(p') \times [\check{e}G_s(p-k')\check{e}'^* + \check{e}'^* G_s(p+k)\check{e}]v(p), \tag{23}$$

where the form of the 2D Green's function G_s in the momentum representation is defined by expression (20b); the properties of the 2D spinor v are defined by formulas (12a)–(12c); and e and e' are the polarization vectors of the initial and final photons. The expression for M coincides with the matrix element of the "4D" Compton effect [30] upon a transition to 2D quantities and convolutions and does not contain an explicit field dependence. The cross sections and probabilities do not contain the field dependence either since the integration with respect to $L_2 dp_2'/(2\pi)$ eliminates the $\delta^{(2)}$ function. If we consider cross processes of the type $\gamma\gamma \longrightarrow e^+e^-$ and $\gamma \longrightarrow \gamma e^+e^-$, additional integration with respect to $L_2 dp_2/(2\pi)$ is equivalent, in accordance with Eqs. (6) and (8), to the integration with respect to coordinate X_1 of the center of the packet; i.e.,

$$\frac{L_2}{2\pi}\int dp_2 = \frac{L_2\gamma}{2\pi}\int_0^{L_1} dX_1 = \frac{L_1 L_2}{2\pi}\gamma, \tag{24}$$

where L_1 is the effective normalization length along the axis 1. Thus, the cross sections and probabilities of such processes are linear functions of the field. It should also be noted that "nonsterile" polarization states of a photon in the 2D version of the theory are states with the electric vector in the momentum–field plane, which can be presented in covariant form as

$$e_\alpha = \frac{(k\varepsilon)_\alpha}{\sqrt{k^2}}, \quad e^2 = -1, \tag{25}$$

where ε is the absolutely antisymmetric tensor in the (0, 3) space introduced by us earlier.

It was mentioned in Section 1 that we consider here the diagrams with an even number of vectorial and (or) pseudovectorial vertices. A matrix element of a loop diagram with a fixed arrangement of vectorial and (or) pseudovectorial vertices corresponds to a tensor (pseudotensor)

$$M_{\alpha_1\ldots\alpha_n}(k^{(1)},\ldots,k^{(n)}) = \int d^4x_1 \ldots \int d^4x_n \times \exp\left[i\sum_{j=1}^{n}(k^{(j)}x_j)\right]\mathrm{Tr}[\tilde{\Gamma}^{(1)}_{\alpha_1} S_s(x_1,x_2) \times \tilde{\Gamma}^{(2)}_{\alpha_2} S_s(x_2,x_3)\ldots\tilde{\Gamma}^{(n)}_{\alpha_n} S_s(x_n,x_1)], \tag{26}$$

in which scalar produces in the exponent are four-dimensional, the form of S_s is specified by formulas (20), (20a), and (20b), $k^{(j)}$ are the momenta of external lines of zero-charge particles, and

$$\tilde{\Gamma}^{(j)}_\alpha = (\tilde{\gamma}_\alpha, \tilde{\gamma}_\alpha\tilde{\gamma}^5). \tag{27}$$

The integration over transverse coordinates can be carried out by applying successively the following integral relation:

$$\int d^2z_\perp f(x_\perp, z_\perp)\varphi((x-z)_\perp) f(z_\perp, y_\perp)\varphi((z-y)_\perp) \times \exp[-i(kz)_\perp] = \frac{2\pi}{\gamma} f(x_\perp, y_\perp)\varphi((x-y)_\perp) \times \tilde{\varphi}(x_\perp, y_\perp)\exp\left[-\frac{i}{2}(k(x+y)_\perp)\right], \tag{28}$$

$$\tilde{\varphi}(x_\perp, y_\perp) = \exp\left\{-\frac{k_\perp^2}{2\gamma} + \frac{1}{2}[k_2(x_1-y_1) - k_1(x_2-y_2)]\right\}. \tag{28a}$$

Since the form of $\varphi((x-y)_\perp)$ implies that

$$((x-y)_{1,2})_{\mathrm{eff}} \sim \gamma^{-1/2}$$

the exponent in Eq. (28a) is of the order of $k_\perp/\sqrt{\gamma}$ and $\tilde{\varphi} \approx 1$ in the given approximation in accordance with condition (19b).

Thus, the $2(n-1)$-fold integration with respect to transverse coordinates eliminates $n-1$ factors $\gamma/2\pi$ appearing in the expression for S_s. The integration with respect to the last pair of transverse coordinates leads to the factor

$$(2\pi)^2\delta^{(1,2)}\left(\sum_{j=1}^{n} k^{(j)}\right).$$

Considering that we are left only with factor $\gamma/2\pi$ and integrating with respect to coordinates (0, 3) in formula (26), we obtain

$$M_{\alpha_1\ldots\alpha_n}(k^{(1)},\ldots,k^{(n)}) = 2\pi\gamma\delta\left(\sum_{j=1}^{n} k^{(j)}\right) \times \int d^2q\,\mathrm{Tr}[\tilde{\Gamma}^{(1)}_{\alpha_1} G_s(q)\tilde{\Gamma}^{(2)}_{\alpha_2} G_s(q+k^{(2)})\ldots\tilde{\Gamma}^{(n)}_{\alpha_n} G_s(q-k^{(1)})], \tag{29}$$

where the 4-momentum is conserved. Here and below, notation $\delta(k)$ is used to denote a 4D δ function.

In the case of an odd number of vertices, the addition of the term with reverse circumvention of the loop to expression (29) gives zero irrespective of the form of $\tilde{\Gamma}_\alpha^{(i)}$ (27); i.e., the extended Farry theorem holds in the 2D version as compared to the 4D version. This means that in expression (17c), we must take into account the next terms in the expansion in η (in reciprocal field); as a result, the matrix elements of such diagrams are independent of the field for $B \gg B_0$.

In the case of an even number of vertices, the resultant tensor can be obtained by reducing expression (29) to symmetric form and does not vanish; i.e., the matrix element is a linear function of the field.

It should also be noted that in strong magnetic fields of the order of 10^{18} G, the actual expansion parameter of perturbation theory in QED becomes a function of the field ($\alpha\ln^2(B/B_0)$, $\alpha(B/B_0)$, $\alpha = e^2 = 1/137$); in such cases, one must carry out summation of the series in perturbation theory as was done in [31] for the main types of compact diagrams in QED.

3. PHOTOGENERATION OF AXIONS AND NEUTRINOS AT NUCLEI

The Lagrangian describing the interaction between an axion and an electron has the form

$$\mathcal{L}_a = \frac{ce}{2f}\frac{\partial a}{\partial x^\alpha}(\bar{\Psi}_e\gamma^\alpha\gamma^5\Psi_e), \tag{30a}$$

and the Lagrangian in the electroweak model in the contact approximation is given by

$$\mathcal{L}_\nu = -\frac{G}{\sqrt{2}}[\bar{\Psi}_e\gamma^\alpha(C_V + C_A\gamma^5)\Psi_e] \times [\bar{\Psi}_\nu\gamma_\alpha(1+\gamma^5)\Psi_\nu], \tag{30b}$$

(c_e is a model-dependent constant of the order of unity), while the structural constants can be expressed in terms of the Weinberg angle:

$$C_V^{(e)} = \frac{1}{2} + \sin^2\theta_W, \quad C_A^{(e)} = \frac{1}{2}$$

for electron neutrinos and $C^{(\mu,\tau)} = C^{(e)} - 1$ for μ and τ neutrinos. Taking into account the QED Lagrangian

$$\mathcal{L} = eA^\alpha(\bar{\Psi}_e\gamma_\alpha\Psi_e) \tag{30c}$$

and using the computational algorithm for loop diagrams in a strong magnetic field described in Section 2, we can obtain the following expression for the matrix element of photogeneration of an axion:

$$\langle f|S|i\rangle_a = \frac{e^\alpha e'^{*\alpha'}A^\beta(\kappa)}{(2q_0 2k_0 2k_0')^{1/2}V^{3/2}}M_{\alpha\alpha'\beta}, \tag{31}$$

$$M_{\alpha\alpha'\beta} = \frac{e^3 c_e\gamma}{6\pi f}J_{\alpha\alpha'\beta}(-k, k', \kappa), \tag{32}$$

where $J_{\alpha\alpha'\beta}$ is a tensor symmetrized in photon lines in the (0, 3) subspace:

$$J_{\alpha\alpha'\beta}(k, k', \kappa) = -\frac{i}{\pi}\int d^2p \times\frac{1}{2}\mathrm{Tr}[\tilde{\gamma}^5\tilde{q}G_s(p)\tilde{\gamma}_\alpha G_s(p+k) \times\tilde{\gamma}_{\alpha'}G_s(p+k+k')\tilde{\gamma}_\beta G_s(p-q)] + 2\text{ transpositions}. \tag{33}$$

In formula (31), V is the normalization volume, $A^\beta(\kappa)$ is the Fourier transform of the external field taken into account in the first Born approximation for the external field, which is given by

$$A_\beta(\kappa) = g_{\beta 0}\frac{8\pi^2\delta(\kappa_0)(Ze)}{\boldsymbol{\kappa}^2}, \tag{34}$$

in the case of a Coulomb field, $\kappa = k - k' - q$ is the transferred momentum, k and k' are the momenta of the initial and final photons, and q is the momentum of an axion.

The form of the tensor $J_{\alpha\alpha'\beta}$ in the low-energy approximation in the electron mass, disregarding the small mass of an axion, was defined in [32], where the three-photon decay of an axion was investigated:

$$J_{\alpha\alpha'\beta}(-k, k', \kappa) = -\frac{4q^2}{15m^6}(k\varepsilon)_\alpha(k'\varepsilon)_{\alpha'}(\kappa\varepsilon)_\beta, \tag{35}$$

$$J_{\alpha\alpha'\beta}k^\alpha = J_{\alpha\alpha'\beta}k'^{\alpha'} = J_{\alpha\alpha'\beta}\kappa^\beta = 0, \tag{35a}$$

where

$$q^2 \equiv q_0^2 - q_3^2 = q_\perp^2$$

in view of the 2D nature of convolutions.

Defining nonsterile polarization states of photons by formula (25) and taking into account relations (14) and (34), we obtain

$$\frac{(k\varepsilon)^\alpha}{\sqrt{k^2}}\frac{(k'\varepsilon)^{\alpha'}}{\sqrt{k'^2}}J_{\alpha\alpha'0}(-k, k', \kappa) = \frac{4q_\perp^2}{15m^6}k_\perp k'_\perp\kappa_3. \tag{35b}$$

In this case, the probability of photogeneration of an axion at a nucleus per unit time, which is determined by the conventional methods, can be written in the form

$$W_a = \left(\frac{8\alpha^2 Zc_e\gamma}{45fm^6}\right)^2\frac{k_\perp^2}{2k_0(2\pi)^5V}\int\frac{d^3k'}{2k_0'}k_\perp'^2 I_a, \tag{36}$$

$$I_a = \int\frac{d^3q}{2q_0}\delta(k_0 - k_0' - q_0)\frac{q_\perp^4(k_3 - k_3' - q_3)^2}{(\mathbf{k} - \mathbf{k}' - \mathbf{q})^4}. \tag{37}$$

Further, we will be interested in the axion emission intensity from unit volume for the concentration n_0 of

nuclei and in accordance with the Planck distribution of the equilibrium emission field, which has a physical meaning. It can easily be seen that this quantity has the form

$$S_a = \left(\frac{8\alpha^2 Z c_e \gamma}{45 f m^6}\right)^2 \frac{n_0}{(2\pi)^8}\int \frac{d^3k}{2k_0}\frac{k_\perp^2}{e^{k_0/T}-1} \times \int\limits_{(k_0' \le k_0)} \frac{d^3k'}{2k_0'} k_\perp'^2 (k_0 - k_0') I_a. \tag{38}$$

We disregard the mass parameter (1) in view of its smallness. For this reason, the temperature dependence can be factorized after corresponding substitutions, and the final result can be written in the form

$$S_a = \frac{691(2\pi)^7}{728\times 45^3}\frac{\alpha^4 Z^2 c_e^2}{(f/m)^2}\left(\frac{T}{m}\right)^{12}\left(\frac{B}{B_0}\right)^2 (n_0 m^2)\tilde{I}_a, \tag{39}$$

where the quantity $\tilde{I}_a$ is a number presented in integral form:

$$\tilde{I}_a = \frac{1}{2\pi}\int_0^1 d\varepsilon \varepsilon^3 (1-\varepsilon)^6 \int_{-1}^1 dx (1-x^2)\int_{-1}^1 dy (1-y^2) \times \int_0^{2\pi} d\varphi \int_{-1}^1 dz (1-z^2)^2 \frac{n_3^2 \tilde{n}^2}{[\tilde{n}^4 - n_\perp^2 (1-\varepsilon)^2 (1-z^2)]^{3/2}}, \tag{40}$$

$$\begin{cases} n_3 = x - \varepsilon y - z(1-\varepsilon), \\ n_\perp^2 = 1 - x^2 + \varepsilon^2 (1-y^2) - 2\varepsilon\sqrt{(1-x^2)(1-y^2)}\cos\varphi, \\ \tilde{n}^2 = 1 - \varepsilon + \varepsilon^2 - \varepsilon x y \\ \quad - \varepsilon\sqrt{(1-x^2)(1-y^2)}\cos\varphi - (x - \varepsilon y)(1-\varepsilon) z. \end{cases} \tag{40a}$$

The dimensionless integration variables correspond to the following variable in the momentum space:

$$z = \frac{q_3}{k_0 - k_0'}, \quad \varepsilon = \frac{k_0'}{k_0}, \qquad x = \cos\theta_{\mathbf{k}}, \quad y = \cos\theta_{\mathbf{k}'}, \quad \varphi = \widehat{\mathbf{k}_\perp \mathbf{k}_\perp'}. \tag{40b}$$

Numerical integration gives $\tilde{I}_a = 0.0046$.

The matrix element of the process of radiation-induced photogeneration of a neutrino pair at a Coulomb center with the Fourier transform $A^\beta(\kappa)$ can be obtained by using expression (30b) and (30c) and has the form

$$\langle f|S|i\rangle_\nu = i\frac{e^\alpha e'^{*\alpha'} A^\beta(\kappa)\mathcal{F}^{\beta'}}{(2k_0 2k_0' 2q_0 2q_0')^{1/2} V^2} M_{\alpha\alpha'\beta\beta'}, \tag{41}$$

where

$$\mathcal{F}^{\beta'} = [\bar{u}_\nu(q)\gamma^{\beta'}(1+\gamma^5)u_\nu(-q')], \tag{42}$$

$$M_{\alpha\alpha'\beta\beta'} = \frac{e^3 G\gamma}{3\sqrt{2}\pi} J_{\alpha\alpha'\beta\beta'}(-k, k', \kappa), \tag{43}$$

where the 2D tensor $J_{\alpha\alpha'\beta\beta'}$ symmetrized in photon lines is given by

$$J_{\alpha\alpha'\beta\beta'}(k, k', \kappa) = \frac{i}{\pi}\int d^2 p \frac{1}{2}\mathrm{Tr}[\tilde{\gamma}_{\beta'}(C_V + C_A\tilde{\gamma}^5) \times G_s(p)\tilde{\gamma}_\alpha G_s(p+k)\tilde{\gamma}_{\alpha'} G_s(p+k+k') \times \tilde{\gamma}_\beta G_s(p-Q)] + 2 \text{ transpositions}, \tag{44}$$

$Q = q + q'$ being the total momentum of the neutrino.

Calculations similar to those made in [32] give, in the low-energy approximation, the following expression:

$$J_{\alpha\alpha'\beta\beta'}(-k, k', \kappa) = \frac{4}{15 m^6}(k\varepsilon)_\alpha (k'\varepsilon)_{\alpha'}(\kappa\varepsilon)_\beta \times [C_V (Q\varepsilon)_{\beta'} + C_A Q_{\beta'}] \tag{45}$$

with properties in α, α', and β indices similar to those of expressions (35a) and (35b). Taking into account the value of the integral over neutrino momenta ($m_\nu = 0$),

$$\int \mathcal{F}^\nu \mathcal{F}^{*\mu}\delta(Q - q - q')\frac{d^3q}{2q_0}\frac{d^3q'}{2q_0'} = \frac{4\pi}{3}[Q^\mu Q^\nu - (Q^2 - Q_\perp^2)g^{\mu\nu}],$$

and form (34) of $A^\beta(\kappa)$, we can present the probability of photogeneration of a neutrino pair at a nuclei per unit time in the form

$$W_\nu = \left(\frac{8\sqrt{2}\alpha^2 Z G\gamma}{45 m^6}\right)^2 \frac{k_\perp^2}{3k_0 (2\pi)^7 V} \times \int \frac{d^3k'}{2k_0'} k_\perp'^2 [I_\nu^{(V)} C_V^2 + I_\nu^{(A)} C_A^2], \tag{46}$$

$$\begin{Bmatrix} I_\nu^{(V)} \\ I_\nu^{(A)} \end{Bmatrix} = \int\limits_{(Q_0^2 - \mathbf{Q}^2) \ge 0} d^3 Q \frac{(k_3 - k_3' - Q_3)^2}{(\mathbf{k} - \mathbf{k}' - \mathbf{Q})^4} \times (Q_0^2 - Q_3^2)\begin{Bmatrix} Q_0^2 - \mathbf{Q}^2 \\ Q_\perp^2 \end{Bmatrix}, \tag{47}$$

while the emission intensity has the form

$$S_\nu = \left(\frac{8\sqrt{2}\alpha^2 ZG\gamma}{45m^6}\right)^2 \frac{2n_0}{3(2\pi)^{10}}\int\frac{d^3k}{2k_0}\frac{k_\perp^2}{e^{k_0/T}-1} \times \int\limits_{(k_0'\le k_0)}\frac{d^3k'}{2k_0'}k_\perp'^2(k_0-k_0')[I_\nu^{(V)}C_V^2 + I_\nu^{(A)}C_A^2]. \quad (48)$$

A result similar to formula (39) has the form

$$S_\nu = \frac{2(2\pi)^7}{135^2}\alpha^4 Z^2 (Gm^2)^2\left(\frac{B}{B_0}\right)^2\left(\frac{T}{m}\right)^{14} \times (n_0 m^2)[\tilde{I}_\nu^{(V)}C_V^2 + \tilde{I}_\nu^{(A)}C_A^2] \quad (49)$$

with the integral representations of the coefficients

$$\tilde{I}_\nu^{(V)} = \prod\left\{(1-\varepsilon)^2(1-z^2) - n_3^2 - n_\perp^2 + [f^2 + 4n_3^2n_\perp^2]^{-1/2}[(n_3^2+n_\perp^2)^2 + 2(1-\varepsilon)^2(1-z^2)(n_3^2-n_\perp^2) + (1-\varepsilon)^4(1-z^2)^2] - 2n_3^2\ln\left(\frac{(f^2+4n_\perp^2n_3^2)^{1/2}+f}{2n_3^2}\right)\right\}, \quad (50a)$$

$$\tilde{I}_\nu^{(A)} = \prod\left\{n_\perp^2 + n_3^2 + [f^2 + 4n_3^2n_\perp^2]^{-1/2} \times [-(n_3^2+n_\perp^2)^2 + (1-\varepsilon)^2(1-z^2)(n_\perp^2-3n_3^2)] + 2n_3^2\ln\left(\frac{(f^2+4n_\perp^2n_3^2)+f}{2n_3^2}\right)\right\}, \quad (50b)$$

$$\prod\{A\} = \frac{1}{2\pi}\int_0^1 d\varepsilon\varepsilon^3(1-\varepsilon)^4\int_{-1}^1 dx(1-x^2) \times \int_{-1}^1 dy(1-y^2)\int_0^{2\pi}d\varphi\int_{-1}^1 dz(1-z^2)A, \quad (50c)$$

$$f = (1-\varepsilon)^2(1-z^2) + n_3^2 - n_\perp^2. \quad (50d)$$

The remaining notation of dimensionless variables is the same as in expressions (40a) and (40b) after the substitution $q_3 \longrightarrow Q_3$. As a result of numerical integration, we obtained the following values: $\tilde{I}_\nu^{(V)} \approx 2.2\times10^{-4}$ and $\tilde{I}_\nu^{(A)} \approx 3.2\times10^{-4}$.

4. INELASTIC SCATTERING OF A PHOTON BY A PHOTON

The matrix element of inelastic scattering $\gamma\gamma \longrightarrow \gamma a$ with the generation of an axion in the low-energy approximation in the electron mass can be obtained from the general relations (29), (30a), (30c), (33), and (35); in the case of allowed nonsterile polarization states (25) of photons, it has the form

$$\langle f|S|i\rangle_a = \frac{(2\pi)^4\delta(k+k'-k''-q)}{(2k_0 2k_0' 2k_0'' 2q_0)^{1/2}V^2}M_a, \quad (51)$$

$$M_a = -\frac{4e^3c_e\gamma}{45\sqrt{\pi}fm^6}q_\perp^6 k_\perp k_\perp' k_\perp'', \quad (52)$$

where k and k' are the momenta of the initial photons, and k'' and q are the momenta of final photon and an axion.

The probability of the process per unit time is defined as

$$W_a = \frac{\alpha^3 c_e^2\gamma^2 k_\perp^2 k_\perp'^2}{45^2\pi^2 f^2 m^{12}k_0k_0'V}I, \quad (53)$$

where I is the invariant in subspace (0, 3),

$$I = \frac{1}{\pi}\int\frac{d^3q}{2q_0}\int\frac{d^3k''}{k_0''}\delta(p-k''-q)q_\perp^4 k_\perp''^2, \quad p = k+k'. \quad (54)$$

A quite cumbersome integration leads to

$$I = \frac{1}{280}\left(p^6 - \frac{1}{3}p^4p_\perp^2 - \frac{1}{3}p^2p_\perp^4 + p_\perp^6\right). \quad (55)$$

Dividing Eq. (53) by the factor

$$\frac{k_0k_0' - \mathbf{k}\mathbf{k}'}{Vk_0k_0'}$$

and taking into account expression (55), we obtain the scattering cross section.

In order to find the axion emission power, we must calculate the explicit form of the 2D vector in (0, 3):

$$I_\mu = \frac{1}{\pi}\int\frac{d^3q}{2q_0}\int\frac{d^3k''}{2k_0''}\delta(p-k''-q)q_\perp^4 k_\perp''^2 q_\mu. \quad (56)$$

Calculations give

$$I_\mu = \frac{p_\mu}{1680}(3p^6 + 2p^4p_\perp^2 - 6p^2p_\perp^4 + 6p_\perp^6), \quad \mu = 0, 3. \quad (57)$$

Thus, the emission power from unit volume can be presented in the form of an integral over the phase volume of initial photons in the equilibrium radiation field:

$$S_a = \frac{16\alpha^3 c_e^2 \gamma^2}{45^2 m^{12} f^2 (2\pi)^8} \int \frac{d^3k}{2k_0} \frac{k_\perp^2}{\exp(k_0/T)-1} \times \int \frac{d^3k'}{2k_0'} \frac{k_\perp'^2}{\exp(k_0'/T)-1} I_0. \quad (58)$$

Integration with respect to angular variables with the factorization of the temperature dependence gives

$$S_a = \frac{128}{45^3 \times 105} \frac{\alpha^3 c_e^2 \gamma^2 T^{15}}{m^{12} f^2 (2\pi)^6} \int_0^\infty \frac{d\varepsilon}{e^\varepsilon - 1} \int_0^\infty \frac{d\varepsilon'}{e^{\varepsilon'} - 1} \times \left(\frac{20}{21}\varepsilon^{10}\varepsilon'^3 + \frac{83}{21}\varepsilon^9\varepsilon'^4 + 21\varepsilon^8\varepsilon'^5 + \frac{136}{5}\varepsilon^7\varepsilon'^6 \right). \quad (59)$$

The value of the double integral in formula (59) found as a result of numerical calculations turned out to be approximately equal to 8.7×10^7.

The final result for the emission power can be written in the form

$$S_a \approx 0.058\alpha^3 c_e^2 \left(\frac{B}{B_0}\right)^2 \left(\frac{m}{f}\right)^2 \left(\frac{T}{m}\right)^{15} m^5. \quad (60)$$

In accordance to the results obtained in [25], the corresponding expression for the neutrino emission power can be written in the form

$$S_\nu \approx 10(1.24C_V^2 + 1.20C_A^2) \times \alpha^3 (Gm^2)^2 \left(\frac{B}{B_0}\right)^2 \left(\frac{T}{m}\right)^{17} m^5. \quad (61)$$

5. DISCUSSION

In order to estimate the role of the effects under investigation, we must set preliminarily the range of temperatures and magnetic field induction, in which the contribution of the diagrams considered above exceeds the contribution from tripole diagrams.

The neutrino emission power in the case of photogeneration at nuclei due to the contribution of triangular diagram in fields with induction $B \gg B_0$ in the Fermi scheme is approximately equal to [24][1]

$$S_\nu^{(F)} \approx 0.6\alpha^3 Z^2 (Gm^2)^2 \left(\frac{T}{m}\right)^{10} (n_0 m^2). \quad (62)$$

[1] In [24], the factor $(2\pi)^{-3}$ is omitted in formula (5); the corresponding correction should be introduced in the subsequent formulas.

Assuming that $C_V \approx C_A \approx 1$ for obtaining an estimate on the basis of formula (49), we get

$$\frac{S_\nu(\gamma(Ze) \longrightarrow \gamma(\nu\bar{\nu}))}{S_\nu^{(F)}(\gamma(Ze) \longrightarrow \nu\bar{\nu})} \approx 3.8 \times 10^{-4} \left(\frac{B}{B_0}\right)^2 \left(\frac{T}{m}\right)^4. \quad (63)$$

If $T = 0.1m$, the contribution of the quadripole diagram to the emission power for the maximum admissible value of $B \sim 10^4 B_0$ might be several times larger than the contribution of the tripole diagrams. If we assume that formula (63) also gives a correct order-of-magnitude estimate of the result for temperatures $T \sim m$ typical of early stages of the evolution of supernovas, the process of radiation-induced photogeneration starts dominating even for $B \sim 10^2 B_0$.

The contribution of the triangular diagram to the inelastic process $\gamma\gamma \longrightarrow \nu\bar{\nu}$ in the Fermi scheme in fields with induction $B \gg B_0$ is given by [33]

$$S_\nu^{(F)} \approx 0.13\alpha^2 (Gm^2)^2 \left(\frac{T}{m}\right)^{13} m^5. \quad (64)$$

Comparing this expression with the contribution of the quadripole diagram (61), we again assume that $C_V \approx C_A \approx 1$ and obtain the relative quantity

$$\frac{S_\nu(\gamma\gamma \longrightarrow \gamma(\nu\bar{\nu}))}{S_\nu^{(F)}(\gamma\gamma \longrightarrow \nu\bar{\nu})} \approx 0.8 \left(\frac{B}{B_0}\right)^2 \left(\frac{T}{m}\right)^4; \quad (65)$$

i.e., the possible range of variation of parameters B and T in the region where $S_\nu(\gamma\gamma \longrightarrow \gamma(\nu\bar{\nu}))$ dominates is slightly wider.

Let us now compare the mechanisms of neutrino luminosity under investigation with the traditional mechanism associated with modified URCA process whose power can be estimated by the order of magnitude as

$$S_\nu(\text{URCA}) \sim 10^{27} \left(\frac{\rho}{\rho_0}\right)^{2/3} \left(\frac{T}{m}\right)^8 \text{ erg/(cm}^3\text{ s)}, \quad (66)$$

where $\rho_0 = 2.8 \times 10^{14}$ g/cm^3 is the characteristic nuclear density and ρ is the average density of the star. Result (49) can be presented in a similar form ($C_V \approx C_A \approx 1$):

$$S_\nu(\gamma(Ze) \longrightarrow \gamma(\nu\bar{\nu})) \sim 10^{20} \left(\frac{B}{B_0}\right)^2 \left(\frac{T}{m}\right)^{14} \left(\frac{Z^2}{A}\right) \left(\frac{\rho}{\rho_0}\right) \text{ erg/(cm}^3\text{ s)}. \quad (67)$$

Assuming that estimate (66) is valid for $T \sim m$ and taking into account the mean value of Z^2/A in the range 1–10 [34], we find that the contribution from the mechanism of radiation-induced photogeneration of neutrinos

at nuclei competes with the URCA process for $B \sim (10^3–10^4)B_0$. This is also in accord with the comments to formula (63).

The emission power defined by formula (61) has the order of magnitude

$$S_\nu(\gamma\gamma \longrightarrow \gamma(\nu\bar{\nu})) \sim 10^{17}\left(\frac{B}{B_0}\right)^2\left(\frac{T}{m}\right)^7 \text{ erg/(cm}^3\text{ s)} \quad (68)$$

and does not compete with the power of the URCA process even for $B \sim 10^4 B_0$, $T \sim m$.

Thus, for the above values of parameters B and T for which the process of radiation-induced photogeneration may become predominant, the lower boundary of the scale of the PQ symmetry breaking and the upper boundary of axion masses can be obtained from the condition

$$S_a(\gamma(Ze) \longrightarrow \gamma a) \lesssim S_\nu(\gamma(Ze) \longrightarrow \gamma(\nu\bar{\nu})). \quad (69)$$

Taking into account formulas (3), (39), and (49), we obtain

$$f \gtrsim 1.7 \times 10^8 \frac{c_e(m/T)}{(0.5C_V^2 + 0.7C_A^2)^{1/2}} \text{ GeV}, \quad (70)$$

$$m_a \lesssim 3.5 \times 10^{-2} \text{ eV}\left(\frac{T}{m}\right)\frac{(0.5C_V^2 + 0.7C_A^2)^{1/2}}{c_e}. \quad (71)$$

Under the assumed values of $T \sim m$, $B \sim (10^3–10^4)B_0$, and for $c_e \sim 1$, this does not contradict the possible range (4) obtained from different considerations.

REFERENCES

1. G. G. Raffelt, *Stars as Laboratories for Fundamental Physics* (Chicago Univ. Press, Chicago, 1996).
2. G. S. Bisnovatyĭ-Kogan, *Physical Problems of Star Evolution* (Nauka, Moscow, 1989).
3. F. Boehm and P. Vogel, *Physics of Massive Neutrinos* (Cambridge University Press, Cambridge, 1987; Mir, Moscow, 1990).
4. M. B. Voloshin and K. A. Ter-Martirosyan, *Theory of Gauge Interactions of Elementary Particles* (Énergoatomizdat, Moscow, 1984).
5. V. Petrosian, G. Beaudet, and E. E. Salpeter, Phys. Rev. **154**, 1445 (1967).
6. W. Fowler and F. Hoyle, Astrophys. J., Suppl. Ser. **9**, 201 (1964).
7. L. Rosenberg, Phys. Rev. **129**, 2786 (1963).
8. M. Delbruck, Z. Phys. **84**, 144 (1933).
9. H. E. Jackson and K. J. Wetzel, Phys. Rev. Lett. **22**, 1008 (1969).
10. M. Bolsterly, Phys. Rev. **94**, 282 (1962).
11. G. Jarlskog, L. Jonsson, S. Prunster, *et al.*, Phys. Rev. D **8**, 3813 (1973).
12. G. S. Bisnovatyĭ-Kogan and S. G. Moiseenko, Astron. Zh. **69**, 563 (1992) [Sov. Astron. **36**, 285 (1992)]; R. C. Duncan and C. Thompson, Astrophys. J. Lett. **392**, L9 (1992); C. Thompson and R. C. Duncan, Mon. Not. R. Astron. Soc. **275**, 255 (1995).
13. D. V. Gal'tsov and N. S. Nikitina, Zh. Éksp. Teor. Fiz. **62**, 2009 (1972) [Sov. Phys. JETP **35**, 1047 (1972)].
14. V. V. Skobelev, Zh. Éksp. Teor. Fiz. **71**, 1263 (1976) [Sov. Phys. JETP **44**, 660 (1976)].
15. V. N. Baĭer and V. M. Katkov, Dokl. Akad. Nauk SSSR **171**, 313 (1966) [Sov. Phys. Dokl. **11**, 947 (1966)].
16. V. I. Ritus, Tr. Fiz. Inst. Akad. Nauk SSSR **111**, 96 (1979); Zh. Éksp. Teor. Fiz. **56**, 986 (1969) [Sov. Phys. JETP **29**, 532 (1969)].
17. A. D. Kaminker, K. P. Levenvish, and D. G. Yakovlev, Pis'ma Astron. Zh. **17**, 1090 (1991) [Sov. Astron. Lett. **17**, 450 (1991)].
18. V. V. Skobelev, Zh. Éksp. Teor. Fiz. **107**, 322 (1995) [JETP **80**, 170 (1995)].
19. Particle Data Group (R. M. Barnett *et al.*), Phys. Rev. D **54**, 1 (1996).
20. V. V. Skobelev, Zh. Éksp. Teor. Fiz. **72**, 1298 (1977) [Sov. Phys. JETP **45**, 682 (1977)]; V. V. Skobelev, Izv. Vyssh. Uchebn. Zaved., Fiz., No. 10, 142 (1975).
21. A. V. Kuznetsov and N. V. Mikheev, Zh. Éksp. Teor. Fiz. **118**, 863 (2000) [JETP **91**, 748 (2000)].
22. Yu. M. Loskutov and V. V. Skobelev, Yad. Fiz. **43**, 1495 (1986) [Sov. J. Nucl. Phys. **43**, 964 (1986)].
23. V. V. Skobelev, Zh. Éksp. Teor. Fiz. **117**, 1059 (2000) [JETP **90**, 919 (2000)].
24. Yu. M. Loskutov and V. V. Skobelev, Yad. Fiz. **31**, 1279 (1980) [Sov. J. Nucl. Phys. **31**, 661 (1980)].
25. Yu. M. Loskutov and V. V. Skobelev, Teor. Mat. Fiz. **70**, 303 (1987).
26. Yu. M. Loskutov and V. V. Skobelev, Teor. Mat. Fiz. **84**, 314 (1990).
27. R. D. Peccei and H. R. Quinn, Phys. Rev. Lett. **38**, 1440 (1977); R. D. Peccei, *CP Violation*, Ed. by C. Jarlskog (World Sci., Singapore, 1989).
28. G. G. Raffelt, Phys. Rep. **198**, 1 (1990).
29. N. P. Klepikov, Zh. Éksp. Teor. Fiz. **26**, 19 (1954).
30. V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, in *Relativistic Quantum Theory* (Nauka, Moscow, 1968; Pergamon, Oxford, 1971), Part 1, p. 388.
31. Yu. M. Loskutov, B. A. Lysov, and V. V. Skobelev, Teor. Mat. Fiz. **53**, 469 (1982); Yu. M. Loskutov and V. V. Skobelev, Vestn. Mosk. Univ., Ser. 3: Fiz., Astron., No. 6, 95 (1983); Vestn. Mosk. Univ., Ser. 3: Fiz., Astron., No. 1, 70 (1984).
32. V. V. Skobelev, Zh. Éksp. Teor. Fiz. **116**, 26 (1999) [JETP **89**, 13 (1999)]; Zh. Éksp. Teor. Fiz. **116**, 2271 (1999) [JETP **89**, 1200 (1999)].
33. Yu. M. Loskutov and V. V. Skobelev, Vestn. Mosk. Univ., Ser. 3: Fiz., Astron., No. 4, 10 (1981).
34. J. W. Negele and D. Vautherin, Nucl. Phys. A **207**, 298 (1973).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 93, No. 4, 2001, pp. 695–705.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 797–809.

ATOMS, SPECTRA, RADIATION

Bremsstrahlung in a Strong Laser Field

A. A. Balakin* and G. M. Fraiman

Institute of Applied Physics, Russian Academy of Sciences, ul. Ul'yanova 46, Nizhni Novgorod, 603600 Russia
**e-mail: abal@appl.sci-nnov.ru*

Received January 31, 2001

Abstract—We consider the bremsstrahlung of electrons as they collide with ions in a strong laser field. The bremsstrahlung spectrum has been found to be enriched in sufficiently strong fields. Particular attention is given to the coherent bremsstrahlung component. We propose a qualitative model that explains our results. The possibility of experimentally observing the coherent bremsstrahlung component in a strong field is discussed.

1. INTRODUCTION

The traditional interest in electron-ion pair collisions in strong electromagnetic fields observed for about 40 years [1–10] has become particularly active in recent years in connection with the invention of powerful short-pulse lasers with energy densities up to 10^{18}–10^{21} W/cm^2, which correspond to ultrarelativistic electron oscillation energies.

The physical phenomena that arise in this case when studying the interaction of such fields with ordinary and cluster plasmas and the possible applications (from laser controlled thermonuclear fusion to the transformation of optical radiation into X-ray radiation) have turned the problem of pair collisions in plasma from a purely academic one to an important applied problem.

Theoretical studies of electron-ion collisions are customarily carried out on the basis of three models. All these models are based on the pair-collision approximation in the plasma theory; i.e., it is assumed that all parameters of the collision integral for a single-particle distribution function can be determined by solving the problem of the scattering of a beam of noninteracting electrons by a single ion. The small-angle scattering model [3–7] gained widespread acceptance; in this model, a rectilinear trajectory is taken as the unperturbed electron trajectory, and all effects are estimated using the theory of perturbations along this trajectory. Clearly, in this approximation, electrons collide with ions at uncorrelated times. More specifically, for a beam with an initially uniform distribution in field phase, the distribution of collision times in field phase will also be uniform. In addition, this approximation disregards the possible attraction of the electron to the ion during scattering; i.e., the electron cannot greatly change its impact parameter. A quantum modification of this model (Born approximation [7]) gives no quantum corrections, because only the first order is retained in the semiclassical expansion.

Another model (low-frequency approximation [8–18]) describes strong collisions with large scattering angles. In this case, an external, sufficiently strong electric field is assumed to accelerate the electron before and after collision, and only the statistic field of the nearest ion is significant during the instantaneous scattering. As in the small-angle scattering model, the collisions are assumed to occur at random times.

All the above approximations yielded results that differed only by a logarithmic factor. The assumptions that the collisions were uncorrelated and that the electron could not change its impact parameter as it oscillated near the ion appear to have been mainly responsible for this difference. This has recently been shown particularly clearly by Silin [4], who obtained the same results as those in the papers cited above directly from the Landau collision integral again with a logarithmic accuracy. The close agreement between the results yielded by the three different (at first glance) approximations was apparently the reason why interest in this subject has waned for more than 30 years.

The importance of taking into account the change in impact parameter during oscillations can be understood from the following considerations. Let an electron with an impact parameter ρ impinge on an ion along the electric field of a wave. Flying past the ion at this distance with a large velocity,

$$v_{\mathrm{osc}} = \frac{eE}{m\omega},$$

it is scattered through a small angle,

$$\delta\theta \sim \frac{2e^2Z}{m\rho v_{\mathrm{osc}}^2} \equiv \frac{2b_{\mathrm{osc}}}{\rho} \ll 1,$$

where e and eZ are the electron and ion charges, respectively; and

$$b_{\mathrm{osc}} = e^2Z/mv_{\mathrm{osc}}^2$$

is the Rutherford radius estimated from the oscillation velocity.[1] As a result, the electron acquires a transverse velocity and reduces its impact parameter at the next passage by the ion in half the laser-field period by

$$(\pi/\omega)v_{osc}\delta\theta = 2\pi r_{osc}b_{osc}/\rho \gg b_{osc},$$

where $r_{osc} = eE/m\omega^2$ is the electron oscillation radius. In particular, if this transverse displacement is equal to the initial impact parameter,

$$\sqrt{2\pi r_{osc}b_{osc}} \equiv \sqrt{2\pi}r_E \gg b_{osc},$$

then the electron suffers a strong scattering with a large change in drift energy. At large electron oscillation velocities and oscillation radii, the corresponding cross section $r_E b_{osc}$ significantly exceeds the scattering cross section, whose value is estimated in all the above models via the square of the Rutherford radius b_{osc}^2 from the oscillation velocity.

As we showed in [11], allowance for multiple electron oscillations near the ion and, as a result, a larger change in the impact parameter during scattering leads to even stronger effects. Thus, the main element of novelty in this and previous papers is as follows. The very concept of strong field for electron scattering by ions is defined not only by the oscillation-to-drift velocity ratio but also by yet another parameter: the ratio of the electron oscillation radius to the size of the region where the ion Coulomb field dominates, r_{osc}/r_E:

$$\frac{v_{osc}}{v} \approx 810\frac{\sqrt{P}}{\sqrt{T}\omega} \approx 130\frac{\sqrt{P}\lambda}{T},$$
$$\frac{r_{osc}}{r_E} \approx 1.3\times 10^4\frac{P^{3/4}}{\omega^2} \approx 3.5\times 10^2 P^{3/4}\lambda^2. \quad (1)$$

Here, the frequency ω is in 10^{15} Hz, the power P is in 10^{18} W/cm^2, the electron temperature T is in electronvolts, and the wavelength λ is in microns. More specifically, if the two parameters given by expression (1) are large, then the fields are said to be strong. If, however, this is not the case, then other regimes take place. If the oscillation velocity is large compared to the drift velocity while the oscillation radius is small compared to r_E (for example, in the case of a high radiation frequency), then an averaged ponderomotive description should be developed. Similarly, if the latter parameter is large while the former parameter is small, then the results are similar to those for a weak field, because the drift in a period significantly exceeds the oscillation radius.

In our previous papers, we focused on the effects attributable to an appreciable change in impact parameter during scattering.[2] Here, we dwell on the subtler effects of such a scattering that were touched upon in [11] only superficially. We will discuss the mechanism of electron separation by phases. The point is that because of the different attraction efficiencies of electrons with different total velocities, which are determined by the field phase at the time when the electron is closest to the ion during each oscillation, electrons can group together. This, in turn, can enhance the energy exchange and increase the effective cross section, as well as produce new and, occasionally, unexpected results. Unfortunately, the analytic model of such attraction is complex. Therefore, we will mainly rely on numerical simulations and simple estimates.

Our numerical analysis of electron–ion collisions shows that correlation effects play an important role during collisions in strong fields [11]. In this case, an oscillating electron repeatedly returns to the same ion while preserving and accumulating the phase memory of the preceding collisions. These correlated returns result, on the one hand, in an increase in the electron attraction by the ion and, on the other, in an electron grouping at certain field phases followed by a strong scattering of the formed electron clusters by the Coulomb center. The increase in attraction must cause the bremsstrahlung to intensify, while the grouping can give rise to coherent radiation from the plasma. Our main goal is to investigate the peculiarities of the bremsstrahlung spectra.

All direct calculations of the bremsstrahlung spectra for electron–ion collisions were performed in the dipole approximation. This implies that we restrict our analysis to nonrelativistic laser-field intensities. In this case, the pumping field may be considered uniform on scales of the region significant for collisions that do not exceed the electron oscillation radius [11], i.e.,

$$r_{osc}k_0 \ll 1,$$

where $k_0 = \omega_0/c$, ω_0 is the field frequency. As in all the papers cited above, the condition of the pair-collision approximation is assumed to be satisfied; i.e., the plasma is considered to be rarefied. This approach allows the main plasma collision parameters to be defined in terms of data on the scattering of a beam of noninteracting electrons oscillating in a strong field by a single ion. In general, unless otherwise specified, we will consider plasma that is transparent to laser radiation ($\omega_0 \gg \omega_p$).

This paper is structured as follows: First, we introduce general expressions for the parameters of the incoherent and coherent radiations during electron–ion collisions in a strong laser field (Section 2). In Section 3, we present our numerical simulations of these parame-

[1] Recall that the impact parameter at which the electron impinging on an ion with this velocity is scattered through an angle of $\pi/2$ is commonly called the Rutherford radius for a given velocity v.

[2] The idea of the importance of allowing for multiple returns of an electron to the same ion has been recently developed in [15] when calculating the multiple ionization probability.

ters and show that the incoherent radiation intensity significantly increases compared to plasma radiation of the same temperature without an external field. Here, we also present spectra for the coherent collisional-current component. In conclusion, we qualitatively discuss the main peculiarities of the spectra. In particular, we show that both the increase in incoherent intensity and the major contribution to the radiation at laser-field harmonics are determined by a small fraction of the electrons that are mostly scattered backward (relative to the laser field) and that, as a consequence, significantly change the drift velocity. Here, we also estimate the applicability conditions for our results and interpret the possible peculiarities of an experimental detection of the effects under study.

2. BASIC RELATIONS

If the pair-collision approximation is used to determine the parameters of electron scattering by an ion, then it will suffice to consider the scattering by an ion of an individual electron (Fig. 1) with charge e and drift velocity $\mathbf{v}_-$ in the strong uniform electric field of a wave polarized along the z axis,

$$\mathbf{E}(\mathbf{R}, t) = E\cos(\omega_0 t - \mathbf{k}_0 \cdot \mathbf{R})\mathbf{z}_0, \tag{2}$$

in terms of the classical equation of motion. We restrict our analysis to the nonrelativistic limit. In this case, the electron radiation may be assumed to be a dipole one. In addition, when considering the motion of a nonrelativistic electron, in view of the condition $k_0 r_{\mathrm{osc}} \ll 1$, the wave field may be assumed to be uniform on collision scales and the coordinate of the nearest ion may be substituted for the electron coordinate R in (2). The equation of motion then takes the form

$$m\ddot{\mathbf{R}} = -\frac{Ze^2\mathbf{R}}{R^3} + \mathbf{z}_0 eE\cos\omega t. \tag{3}$$

Equation (3) has the following characteristic scales:

$$r_E = \sqrt{\frac{eZ}{E}} = \sqrt{b_{\mathrm{osc}} r_{\mathrm{osc}}}, \tag{4}$$

$$\omega_E = \sqrt[4]{\frac{eE^3}{m^2 Z}}, \quad v_E = \sqrt[4]{\frac{Ze^3 E}{m^2}}. \tag{5}$$

In dimensionless variables with scales (4), the equation of motion is (the notation is the same)

$$\ddot{\mathbf{R}} = -\frac{\mathbf{R}}{R^3} + \mathbf{z}_0\cos\Omega t. \tag{6}$$

As in [11], it is convenient to pass to an oscillating coordinate system:

$$\mathbf{r} = \mathbf{R} + \frac{\mathbf{z}_0}{\Omega^2}\cos\Omega t. \tag{7}$$

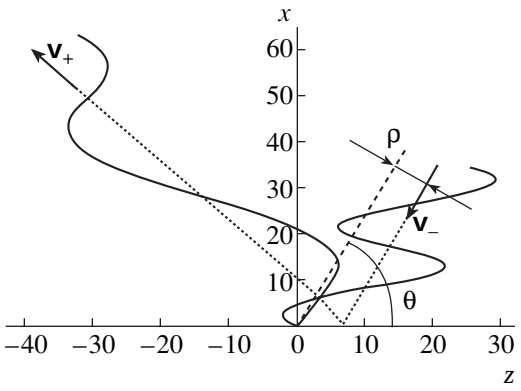

Fig. 1. A typical electron trajectory (solid line), the drift trajectory (dotted line), and parameters of the scattered electron (ρ is the impact parameter, and θ is the angle between the velocity and field E).

The equation of motion for the electron drift coordinate $\mathbf{r}$ (7) then takes the form

$$\ddot{\mathbf{r}} = \nabla\frac{1}{\left|\mathbf{r} - \dfrac{\mathbf{z}_0}{\Omega^2}\cos\Omega t\right|}. \tag{8}$$

In the absence of a Coulomb potential, the drift trajectory is a straight line. Of course, this trajectory can become very complex in a periodically oscillating potential [11]; hence the effects analyzed in the paper.

Equations (6) and (8) include only one dimensionless parameter, the dimensionless frequency

$$\Omega = \omega_4\sqrt{\frac{m^2 Z}{eE^3}} = \frac{\omega_0}{\omega_E} = \sqrt[4]{\frac{b_{\mathrm{osc}}}{r_{\mathrm{osc}}}} = \sqrt{\frac{r_E}{r_{\mathrm{osc}}}}. \tag{9}$$

Another parameter of the problem is the ratio of the initial drift velocity of the scattered electron to v_E from (4).

According to (1), the region in which the following conditions are simultaneously satisfied is below called the region of strong field. (i) The electron drift velocity is small compared to the electron oscillation velocity. (ii) The field amplitude is sufficiently large [11], i.e., in view of (9),

$$\frac{v}{v_{\mathrm{osc}}} = \Omega v \ll 1, \quad \frac{\omega}{\omega_E} = \Omega \ll 1. \tag{10}$$

The first condition implies that an electron can make multiple oscillations when moving near an ion in the wave field. The second condition is the condition for the size of the region with a significant ion Coulomb field being small compared to the electron oscillation amplitude.

Let us determine the electron radiation during collisions in the above approximations. In the absence of a laser field, the effective radiation, which is estimated as the spectral density of the incoherent radiation from the

beam of electrons scattered by a single ion [13], is customarily analyzed in problems. We use a natural generalization of this characteristic by considering the radiation produced by the particle drift, i.e., the spectral density of the incoherent radiation from a beam of oscillating electrons with a given initial drift velocity:

$$\chi_\omega = \frac{2e^2\omega^4}{3c^3}\int \langle |\mathbf{r}_\omega|^2\rangle_\varphi d^2\boldsymbol{\rho}, \tag{11}$$

where

$$\mathbf{r}_\omega = \int_{-\infty}^{\infty} \mathbf{r}(\tau)e^{i\omega\tau}d\tau$$

is the Fourier spectrum of the dipole moment for an individual electron; φ is the entrance phase, which is defined as the field phase at the time when the electron passes the plane of impact parameters; in what follows, $\langle\ldots\rangle_\varphi$ denotes averaging over the entrance phase. The plane of impact parameters $\boldsymbol{\rho}$ (Fig. 1) is perpendicular to the initial electron velocity vector and lies so far from the ion that the effect of the Coulomb field on the motion in this region may be ignored.

In particular, using this parameter for plasma with electron density n_e and ion density n_i, we have for the spectral density of the incoherent radiation from a unit plasma volume

$$dI_{\text{inc}} = \chi_\omega n_e n_i v_T d\omega. \tag{12}$$

For numerical simulations, it is convenient to introduce a dimensionless quantity χ_{inc}:

$$\chi_{\text{inc}}(\omega) = \chi_\omega \frac{3c^3}{2e^2 r_E^4 \omega_E^2}. \tag{13}$$

As a parameter of the coherent radiation, we may choose the coherent cross section χ_{coh}, which is proportional to the coherent collisional-current density. Let us write out an expression for the amplitude of the nth harmonic for this current (see the Appendix):

$$\begin{aligned} \mathbf{j}_n^{\text{col}} &= en_e n_i v_- \frac{2\pi}{\omega}\boldsymbol{\chi}_{\text{coh}}(n\omega_0), \\ \boldsymbol{\chi}_{\text{coh}}(\omega) &= \int \omega^2 \langle \mathbf{r}_\omega\rangle_\varphi d\boldsymbol{\rho}. \end{aligned} \tag{14}$$

The introduced quantity χ_{coh} has the dimensions of area multiplied by velocity, and all parameters of the coherent radiation are expressed in terms of this quantity. As is clear from (14), the collisional-current density is proportional to the product of the electron and ion densities. Qualitatively, this is explained by the fact that the electron current itself is proportional to the product of the electron density by the change in electron velocity averaged over the initial conditions, which, in turn, is proportional to the collision frequency and, consequently, to the ion density.

Note that, in contrast to the incoherent scattering cross section (11), the introduced coherent scattering cross section χ_{coh} at pumping-field harmonics is complex-valued. The phase of this quantity corresponds to the phase shift between the pumping field and the coherent harmonic of the collisional current. Similarly, it is convenient to introduce a dimensionless quantity $\chi_{\text{coh}}^{\text{new}}$ for calculations:

$$\chi_{\text{coh}}^{\text{new}} = \frac{\chi_{\text{coh}}}{r_E^3\omega_E}.$$

In what follows (as previously), the subscript "new" is omitted.

Below, we focus on calculating the spectral parameters χ_{inc} (13) and χ_{coh} (14).

3. SPECTRAL PARAMETERS

To determine the spectral parameters, we calculated the bremsstrahlung cross sections given by relations (11) and (14) for the coherent and incoherent cases. During our numerical experiment, we computed the dipole moment of electron beams with drift velocities from $0.03v_{\text{osc}}$ to $3v_{\text{osc}}$ and with impact parameters up to $5r_{\text{osc}}$. The computations were performed for five frequencies, Ω = 0.1, 0.2, 0.32, 1, and 3.2. The electrons were assumed to be initially uniformly distributed in entrance phase. The initial distance from the Coulomb singularity to the leading particle center was chosen to be $3r_{\text{osc}}$; a particle was considered to have left the interaction region when the leading center receded from the Coulomb singularity by more than $3r_{\text{osc}}$. Particular attention was given to the choice of a sufficient number of particles in the field period. The computations were performed for the longitudinal ($\mathbf{v}_- \parallel \mathbf{E}$) and transverse ($\mathbf{v}_- \perp \mathbf{E}$) scattered beams. We computed the bremsstrahlung during electron scattering by an ion up to the 20th harmonic of the laser-pumping field.

As was shown in [11], the main difference of collisions in strong fields is the appearance of a sizable fraction of electrons scattered through large angles. Their appearance is associated with the strong attraction of the electron due to its multiple correlated oscillations about the ion (Fig. 1). Because of this adiabatic pulling (we call it the parachute effect), the electrons fall within a close vicinity of the ion and are subsequently scattered through large angles.

In Fig. 2, the collision energy and phase are plotted against the impact parameter and initial phase of the scattered electrons. Clearly, these dependences are periodic in initial phase, because the pumping field is periodic. Note an important peculiarity of the dependence of the particle drift energy on impact parameter, $\Delta w(\rho)$. In strong fields, an electron can greatly change its energy due to the parachute effect (darker shade in panel (Fig. 2a)) even if its initial impact parameter was

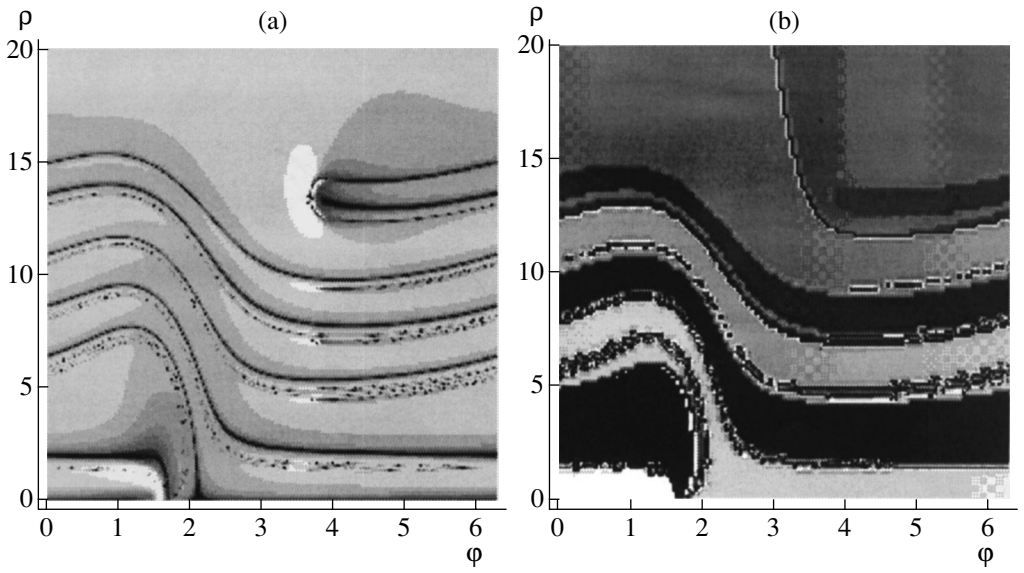

Fig. 2. Collision energy (a) and phase (b) versus impact parameter and initial entrance phase of the scattered electrons for $v = 1$ and $\Omega = 0.1$ (longitudinal scattering, $\mathbf{v}_- \parallel \mathbf{E}$).

much larger than the Rutherford radius b_{osc} estimated from the particle oscillation velocity. Similarly, Fig. 2b shows the grouping effect. Here, the same shade corresponds to the same phase of the last (closest) electron passage near the ion corresponds. The two effects are seen to take place in the region of a significant change in impact parameter during the scattering.

The boundary of this region can be easily estimated from simple analytic considerations. Indeed, in the oscillating coordinate system (7) and (8), the ion oscillates and the unperturbed electron trajectory is a straight line. In each return of the ion, the drift trajectory slightly changes its slope (instantaneous collision) and approaches the ion. To gain a larger drift energy (to significantly change its longitudinal momentum), the electron must fall within a close vicinity of the ion, such that

$$r \leq b_{osc} \ll b_v \ll r_{osc},$$

where $b_{osc} = e^2Z/mv^2$ is the Rutkerford radius of electron and v is electron velocity.

Let us estimate the maximum impact parameter at which this is possible. We restrict ourselves to longitudinal scattering. Let an electron be scattered by an ion through a small angle

$$\delta\theta \sim b_v/\rho_{att} \ll 1.$$

In that case, it must have time to pass in a close vicinity of the ion for a large change in energy. This is possible if it can come to a distance equal to its initial impact parameter in the time it takes to traverse the ion oscillation radius ρ; i.e., the following condition must be satisfied:

$$\rho_{att}/2r_{osc} \approx \delta\theta$$

or

$$\rho_{att} \approx \sqrt{2b_v r_{osc}} = \frac{\sqrt{2}}{\Omega v}. \quad (15)$$

Thus, the parachute effect increases the area from which particles can be scattered through large angles and, accordingly, can radiate more strongly (Fig. 3).

First, we see in Fig. 3 a significant increase in the spectral energy density of the bremsstrahlung in strong fields. Thus, whereas the estimate

$$\chi \propto \frac{1}{v^2}$$

is valid in conventional models in the plateau region, our results are well fitted by the dependence

$$\chi \propto \frac{\pi}{2\Omega^2 v^2}$$

for longitudinal incidence. We see an increase in the bremsstrahlung with pumping-field intensity rather than its decrease, as in conventional models.

This result can be interpreted as follows. It implies that not all of the plasma electrons but only a relatively small percentage of them scattered through large angles produce the bulk of the bremsstrahlung. Previously [11], we called these electrons representative electrons. Indeed, the average change in their velocity is of the order of the oscillation velocity $1/\Omega$; they are collected from the area

$$\rho_{att}^2 \sim \frac{1}{\Omega^2 v^2}$$

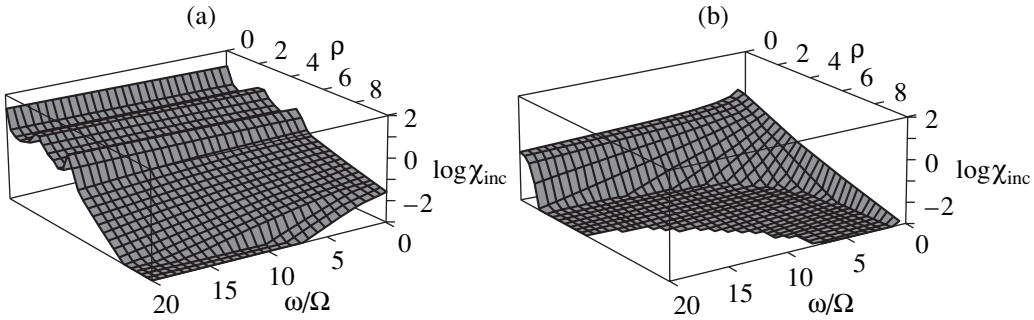

Fig. 3. Spectral energy density of the bremsstrahlung versus impact parameter: (a) numerical simulations and (b) conventional models.

and their fraction is of the order of Ω^2. Their motion [11] is a smooth pulling toward the ion and an instantaneous collision with an abrupt change in particle momentum (and energy). Thus, they radiate energy with a spectral density of the order of

$$\chi \approx \left(\frac{1}{\Omega}\right)^2 \frac{1}{\Omega^2 v^2}\Omega^2 = \frac{1}{\Omega^2 v^2}, \qquad (16)$$

which is comparable to the value that we found by numerical simulations.

Note that, according to [13], the radiation intensity for a particle moving in the way described above must be equal to the square of the change in velocity. For representative electrons, this quantity is approximately equal to the change in particle energy. In other words, the incoherent bremsstrahlung cross section must be equal to the total change in beam energy, as our computations show. For a beam of particles with the initial velocity perpendicular to the electric vector, the change in energy is much smaller [11]. Accordingly, the spectral power density of the radiation from such particles is also weaker:

$$\chi \approx \frac{1}{v^2}. \qquad (17)$$

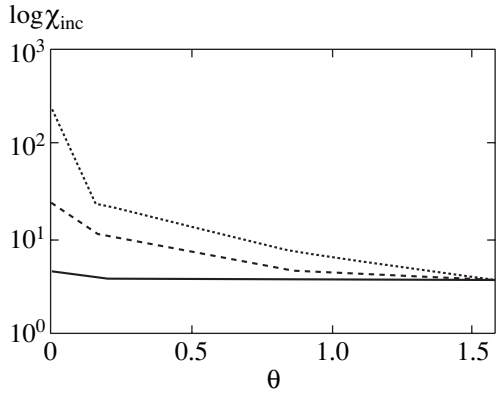

Fig. 4. Incoherent bremsstrahlung cross section versus angle between the initial beam velocity and the field for $v = 1$ at various pumping-wave frequencies: $\Omega = 1$ (weak field), solid line; $\Omega = 0.316$, dashed line; and $\Omega = 0.1$, dotted line.

Integration over the solid angle in the velocity space of impinging particles yields a dependence of the incoherent bremsstrahlung cross section in the form (17). Thus, the formula for conventional models from the weak-field region remains valid, although there is a fraction of the particles that radiate much more strongly. Here, the situation is similar to the situation with energy exchange [11] (Fig. 4).

Figure 2b reflects another important peculiarity of the electron scattering in strong fields. It shows the field phase at the time of electron–ion collision (φ_{col} is the collision phase) taken in the modulus of 2π, the pumping-field period, as a function of their impact parameters and initial phase. The same shade corresponds to the same collision phase. In the figure, we see alternating regions (black and white) with the sharp transitions between them corresponding to representative electrons. This implies that almost all plasma electrons come close to the ion[3] at one of these two pumping-field phases. This phase grouping is prepared by the adiabatic electron drift (parachute effect). Recall that in conventional models [2–10], the collisions are assumed to be uniformly distributed in field phase.

Figure 5 demonstrates the grouping effect in more detail. It shows the particle distribution in collision phase as a function of Ω and v.

It is easy to see that as the v_{osc}/v_T ratio increases, all electrons are collected into two narrow peaks near the maxima of the oscillation velocity. The width of these peaks is proportional to the v_T/v_{osc} ratio. We see from Fig. 5a that at a large thermal electron velocity, $v_T > v_{osc}$, the phase grouping vanishes and the conventional collision model for a weak field becomes applicable.

The width of the peaks in the electron distribution function in collision phase determines the width of the coherent bremsstrahlung spectrum. The narrower the peaks in the distribution function, the wider the coherent bremsstrahlung spectrum (cf. Fig. 6). Thus, the number of peaks increases with v_{osc}/v_T. As expected, the coherent bremsstrahlung vanishes at $v_T > v_{osc}$.

[3] A more detailed analysis indicates that all electrons from the attraction region come within an ion vicinity of size r_E by the collision time.

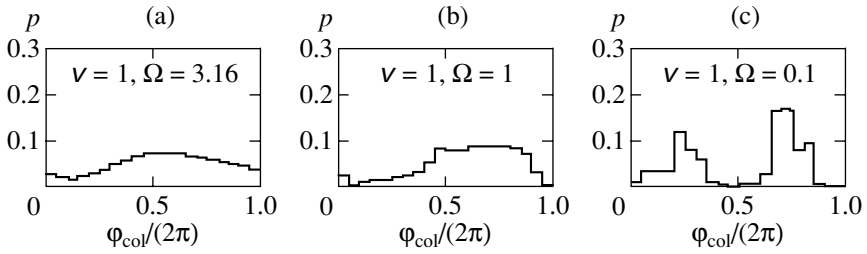

Fig. 5. Particle distribution in collision phase for longitudinal incidence as a function of ν and Ω.

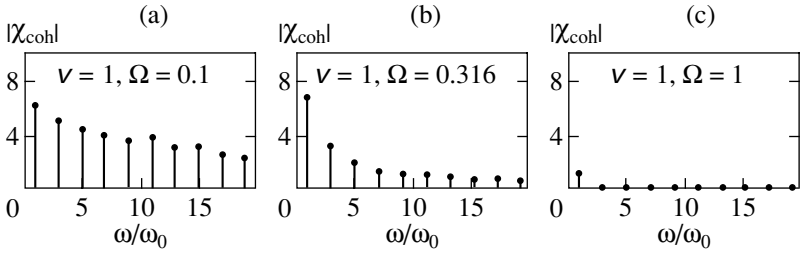

Fig. 6. Coherent bremsstrahlung cross section versus harmonic number for various pumping-field frequencies Ω.

The presence of mostly odd harmonics in the coherent current spectrum stems from the fact that the grouping is realized twice during one half-period at field phases shifted from each other by half the period (Fig. 5). This means that during each grouping peak, the electrons are scattered in antiparallel to the instantaneous oscillation velocity. Thus, the collisional current in each period is represented as two peaks of opposite polarity shifted by half the period.

The results of our numerical simulations for the cross section of coherent bremsstrahlung at the nth harmonic in strong fields can be fitted by the dependence

$$\chi_{\mathrm{coh}} \propto \frac{1}{\Omega \nu^2 n}. \tag{18}$$

This parameter can be estimated [cf. the estimate of the incoherent radiation intensity (16)] from the current of representative electrons as the product of their velocity ν_{osc} by the cross section from which they are collected $\pi\rho_{\mathrm{att}}^2$, by their fraction from the entire number of electrons Ω^2, and by the factor Φ that characterizes the degree of particle grouping in collision phase:

$$\chi \approx \nu_{\mathrm{osc}}\rho_{\mathrm{att}}^2\Omega^2\Phi_n(\Omega\nu) = \frac{\Phi_n(\Omega\nu)}{\Omega\nu^2}. \tag{19}$$

Note that $\Phi(\nu\Omega)$ is bell-shaped: it is equal to unity for $\nu\Omega \ll 1$ and rapidly decreases for $\nu\Omega \geq 1$. Unfortunately, the domain of parameters in our bremsstrahlung computations was chosen in such a way that we could not find a unique velocity dependence of the cross section χ_{coh}, because the parameter $\nu\Omega$ fell within the domain of inflection of $\Phi(\nu\Omega)$, while computations with very low drift velocities are difficult to perform. We give the velocity dependence in (18) as $\propto 1/\Omega\nu^2$, because both simple and more complex models yield values for χ_{coh} that are in good agreement with those obtained in our computations.

In conclusion, two important peculiarities of the coherent radiation (18) should be noted:

(i) The radiation has a wide spectrum that decreases as $1/n$, where n is the harmonic number.

(ii) A small percentage of the representative electrons mainly contribute to the coherent radiation.

4. CONCLUSION

In conclusion, we list those important peculiarities that have been revealed by our numerical simulations.

First, the impact parameter (the minimum distance to the ion in each field period) was found to decrease, because the electron can be attracted to the ion during several successive oscillations about it. Since the electron approaches the ion during each passage and, consequently, the Coulomb field increases, the scattering angle during each subsequent oscillation will be larger than the previous one. It should be noted here that although almost all scatterings when the electron passes by the ion may be considered small-angle ones, the total scattering is not necessarily a small-angle one (primarily for the representative electrons). In other words, when simulating electron-ion collisions, we cannot restrict ourselves to the small-angle approximation [11] but must also take into account the possibility of a large change in angle and small impact parameters during the last collision.

This remark is of particular importance in understanding the bremsstrahlung-related effects. The point is that the contribution of an individual electron to the bremsstrahlung is determined by the change in its momentum rather than in its energy. In particular, at frequencies that are low compared to the collision time, the amplitude and intensity of the radiation field are proportional, respectively, to the change in the momentum itself and the square of this change [13]. That is why the particles with a large change in longitudinal momentum (along the laser field) mainly contribute to the radiation in our case, although their percentage is low. These are the representative electrons. A sizable fraction of these particles are scattered with a large component of the change in momentum (of the order of mv_{osc}) in the direction opposite to the oscillation velocity at the time of the last collision. Given that the incoherent bremsstrahlung cross section in the instantaneous-collision model (at low frequencies), according to [13], is proportional to the sum of the squares of the changes in momenta along and across the pumping field averaged over the field phase and integrated over the impact parameters, we immediately obtain estimates (16), which are in good agreement with our numerical results.

Before analyzing the radiation at harmonics, we make two more useful remarks associated with incoherent-bremsstrahlung studies.

First, the bremsstrahlung spectrum in strong fields corresponds to the white-noise spectrum; i.e., it is essentially independent of frequency, including low frequencies (Fig. 5). This is because the representative electrons responsible for this radiation can emerge only from a bounded region of impact parameters (parachute region). This region has a sharp boundary [12], so the Coulomb-logarithm effect responsible for the logarithmic divergence of the bremsstrahlung spectrum in conventional models vanishes.

The logarithmic divergence in the Landau collision integral and in the bremsstrahlung theory is known to be associated with calculations of the transport collision cross section in plasma. In this case, integration over the impact parameters is restricted to the size of the nonadiabatic-interaction zone, $\rho = v/\omega$. In strong fields (10), the parachute region is definitely larger [11] than this zone and is sharply outlined (the electron cannot be attracted to the ion from large distances to a distance of the order of the Rutherford radius estimated from the oscillation velocity). This is necessary for significant energy exchange. That is why no increase in the bremsstrahlung spectrum is observed at low frequencies. However, for a transverse electron entrance at moderately low velocities (larger than $r_E\omega$), the parachute effect is not observed, and the radiation is the same as in the problem without a field. It determines the total radiation (after integration over the entrance angle). It should be noted that when estimating $\delta v d\rho$ in strong fields, we again derive an expression similar to that obtained without a field. In this case, a large quantity should be substituted for the Coulomb logarithm. Thus, the solution of the problem without a field gives a lower limit for this quantity.

Second, we found the bremsstrahlung spectrum to be virtually independent of the pumping-field polarization (Fig. 4). The reason is that the parachute effect is insensitive to polarization [12], because the oscillatory motion now reduces to circular motion and, being at the smallest distance from the ion, the electron again suffers a scattering. As a result, during the next passage (oscillation), it comes even closer to the ion. The problem of scattering in a circularly polarized field requires a special, more detailed analysis.

The role of representative electrons in the coherent radiation at harmonics is even more significant. The point is that since the problem is isotropic, there is no radiation associated with the change in transverse (relative to the pumping field) momentum in the dipole approximation. Therefore, the coherent radiation is entirely attributable to the change in momentum along the field. In this case, since the electrons that undergo small energy exchange change their longitudinal momentum only in the second order in scattering angle, the representative electrons must mainly contribute to the radiation. Thus, we obtained estimate (19) for the amplitude of the coherent scattering cross section. The first factor v_{osc} corresponds to the characteristic change in longitudinal momentum of the representative electrons; the second factor, to the area of the parachute region; and the third factor, to the fraction of representative electrons. This estimate agrees with our results in the dependence on laser-field intensity and in order of magnitude. Of course, a better estimate can hardly be obtained, because a detailed comparison requires a more accurate model of the grouping effects, the information on which is contained in the function Φ. In estimate (19), however, we restricted ourselves to the limit of the delta approximation; i.e., we assumed all the last scatterings to take place exactly at the phases corresponding to the oscillation-velocity maxima (at the phases of pumping-field zeros). Unfortunately, we have failed to construct a convincing analytic model of the grouping effect (Fig. 5). Therefore, in all our estimates, we took this prerequisite as an experimental result.

As regards the qualitative aspect of the issue, we can only note that, since the scattering angle of the electron during each passage depends on its velocity, which, in turn, depends on the field phase at the time of the electron passage at the closest distance from the ion, the energy-exchange efficiency determined by the field phase at the collision time (the last passage near the ion) depends entirely on the field phase. In essence, this is the mechanism of electron separation in field phase. The result of its action is the arrival of almost all electrons to the ion at certain phases of the external field, i.e., the effect of grouping in phase

It thus follows that the appearance of a weak external high-frequency field does not result in an appreciable modification of the bremsstrahlung cross section at a thermal velocity of the electrons larger than their oscillation velocity. The bremsstrahlung intensity increases in strong fields, because the external high-frequency field, on the one hand, pulls the electrons toward the ion (parachute effect) and, on the other hand, accelerates them to large velocities, compared to the thermal velocity, thereby increasing the acceleration and, accordingly, the particle bremsstrahlung intensity. These two effects combine to produce an increase in the bremsstrahlung intensity in strong fields.

In conclusion, we give estimates for the intensities of the coherent and incoherent radiations and for the applicability conditions for our results in dimensional form. Below, the electron temperature T is in electronvolts, the power P is in 10^{18} W/cm^2, the frequency is in 10^{15} Hz, the wavelength λ is in microns, the density n is in 10^{18} cm^{-3}, the total radiation intensity I_{coh} is in watts, and the pulse duration τ is in 10^{-15} s.

Let us first estimate the characteristic radiation parameters. We begin with the incoherent radiation. We find from formulas (11) and (18) that the incoherent radiation intensity per unit volume is approximately equal to

$$dI_{\mathrm{inc}} \approx 9 \times 10^{-6} \frac{n^2 Z^2}{\sqrt{T}} d\omega. \tag{20}$$

This expression is a lower limit. At low temperatures, $v_T \ll r_E\omega$, the radiation intensity apparently reaches the level

$$dI_{\mathrm{inc}} \approx 0.4 \frac{n^2 P^{3/4}}{\sqrt{T}\omega^2} d\omega. \tag{21}$$

Similarly, we can derive an expression for the lower limit on the coherent radiation at the kth harmonic from formulas (14), (34), and (18):

$$(I_{\mathrm{coh}})_k \approx 2 \times 10^2 \frac{L n^2 \omega Z^4}{k^2 P T} \delta_n\left(\frac{\omega_p a}{c}\right). \tag{22}$$

Here, δ_n is the factor associated with the geometrical beam size [see formula (35) in the Appendix], and L is the length of the strong-field region in microns (Fig. 7). The coefficient of transformation to the kth harmonic can be written as

$$\eta_k \approx 5 \times 10^{-8} \frac{n^2 Z^4}{a k^2 P^2 T \lambda^5}\, \delta_n\left(\frac{\omega_p a}{c}\right), \tag{23}$$

where a is the radius of the strong-field region in microns (Fig. 7). Note that since the coherent component of the collisional-current density is proportional to the square of the density (14), the coherent radiation intensity (34) at first glance must be proportional to the fourth power of the density. Actually, having solved the electrodynamic problem of collisional-current radiation (see the Appendix), we obtained only the second power of the density. To be more precise, the result proves to be even more complex, because the factor δ_n itself is a complex function of the density.

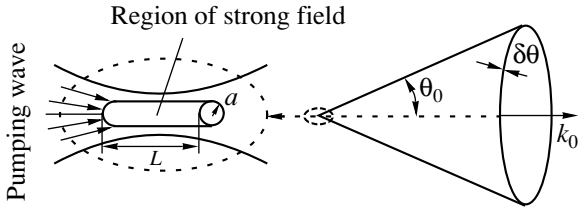

Fig. 7. Admissible angles of the radiated waves.

Let us write out the applicability conditions for our results. Clearly, the strong-field conditions (10) must be satisfied; these can be written in dimensional form as

$$v \ll v_{\mathrm{osc}} \longleftrightarrow T \ll 6.67 \times 10^5 \frac{P}{\omega^2}, \tag{24}$$

$$\omega \ll \omega_E \longleftrightarrow \omega \ll 1.12 \times 10^2 P^{3/8}. \tag{25}$$

Another important condition is the condition imposed on the pulse duration τ_{imp}: it must be longer than the time it takes to traverse the characteristic region of interaction with the ion (of the order of two oscillation radii) with the drift velocity:

$$\tau_{\mathrm{imp}} \gg \frac{v_{\mathrm{osc}}}{v}\frac{2\pi}{\omega} \longleftrightarrow \tau_{\mathrm{imp}} \gg 5.12 \times 10^3 \sqrt{\frac{P}{T}}\frac{1}{\omega^2}. \tag{26}$$

It should be noted that our numerical calculations yielded a sizeable fraction of the captured electrons, the particles that did not leave the interaction region r_{osc} in a time much longer than the time it took to traverse this region, $2r_{\mathrm{osc}}/v_T$. The fraction of such electrons increases with field.

Yet another condition arises from the classical description we used. It can be shown [12] that it is equivalent to the condition

$$r_E \gg \lambda_{\mathrm{osc}} = \frac{h}{m v_{\mathrm{osc}}} \longleftrightarrow P \gg 2.4 \times 10^{-8} \frac{\omega^4}{Z^2}. \tag{27}$$

Thus, the applicability conditions for the classical description in the pair-collision problem become less rigorous with growing field. This can be qualitatively explained by the fact that the energy exchange proper takes place a region of size r_E that is inversely proportional to the square root of the field amplitude (4). Since the de Broglie wavelength of an electron estimated from its oscillation velocity decreases even faster (inversely proportional to the first power of the field), the quantum corrections must be negligible.

The condition imposed on the density is most stringent. Formally, our results are valid only in sufficiently rarefied plasma with $N \ll (\lambda v/c)^{-3}$ (λ is the radiation wavelength in microns). However, one might expect similar effects to take place in denser plasma as well. Indeed, of important is only the fact that no more than one ion is in the volume of the attraction region, $r_{\rm osc}\rho_{\rm att}^2$, because the presence of another ion on the path of the electron scattered by the ion can significantly disrupt the dynamics and can minimize all correlation effects. In dimensional variables, this condition is

$$n \le 3 \times 10^{15} \frac{T\omega^4}{PZ} \approx 4 \times 10^{16} \frac{T}{\lambda^4 PZ}. \quad (28)$$

In denser plasma, the effects described here will show up incompletely (they will be weakened, the dependence on parameters will change), but they will take place up to densities

$$Nb_{\rm osc}^2 r_{\rm osc} \le 1.$$

In such plasma, the neighboring ions do not affect the electron trajectory over the period of change in the external field and allow it to make one oscillation and to be attracted to the ion from distances of the order of

$$r_E = \sqrt{b_{\rm osc} r_{\rm osc}}.$$

We did not perform detailed calculations of this case, but we believe from general considerations that, compared to conventional models, the incoherent bremsstrahlung must be enhanced by a factor of $r_{\rm osc}/b_{\rm osc}$ (by analogy with the increase in effective cross section [11]) and a weak coherent bremsstrahlung must appear. In even denser plasma with

$$Nb_{\rm osc}^3 \sim 1,$$

the correlation effects completely vanish and conventional models become applicable [16].

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project no. 99-02-16443).

APPENDIX

Parameters of Coherent Bremsstrahlung

Let us derive relation (14). Since, as was said above, the electrons group together in phase, it is of interest to study the coherent electron bremsstrahlung. Since the collision times and the external-field phases are correlated, the dipole moment depends not just on time but on

$$t - \mathbf{r}_i \cdot \mathbf{k}_0/\omega_0, \quad (29)$$

where $\mathbf{r}_i$ is the collision coordinate. In other words, a wave of bremsstrahlung flashes running with the pumping-field wave is excited in plasma.

To determine the bremsstrahlung intensity, let us calculate the collisional electron current (we ignore the ion current, because the parameter m_e/M_i is small). In accordance with the general philosophy of the pair-collision approximation, let us first consider the drift current $j_{\rm eff}$ that corresponds to the scattering of an electron beam with density n_e and unperturbed velocity $\mathbf{v}_-$ by a single stationary ion:

$$\mathbf{j}_{\rm eff} = en_e \left\langle \int (\mathbf{v}(\boldsymbol{\rho}, t, t_0) - \mathbf{v}_-) d\boldsymbol{\rho} \right\rangle_\varphi. \quad (30)$$

The quantity defined in this way has the dimensions of current. Averaging over the initial phase φ corresponds to a periodicity of the external field in time. It follows from this periodicity that the current $\mathbf{j}_{\rm eff}$ is also a periodic function and can be represented as a Fourier series with harmonics $\mathbf{j}_n$, such that

$$\mathbf{j}_n = en_e \boldsymbol{\chi}_{\rm coh}(n\omega_0). \quad (31)$$

The quantity $\chi_{\rm coh}$ [see (14)] has the dimensions of area multiplied by velocity; the coherent radiation intensity is expressed in terms of this quantity.

Indeed, using (30) and (31), the collisional-current density in plasma can be written as

$$\mathbf{j}_n^{\rm col} = \frac{2\pi n_i v_-}{\omega} \mathbf{j}_{\rm eff} = \sum_n \mathbf{j}_n^{\rm col} \exp(in\omega_0 t) = \frac{2\pi e n_e n_i v_-}{\omega} \boldsymbol{\chi}_{\rm coh}(n\omega_0). \quad (32)$$

To determine the intensity of the radiation from plasma, we must solve the electrodynamic problem of collisional-current radiation (32) by taking into account the specific geometry of the plasma profile and density. The current-radiation intensity at frequency $n\omega_0$ into a solid angle do can be calculated using the standard solution of the Helmholtz equation by substituting the current $\mathbf{j}_n^{\rm col}$ in it:

$$I_n = \frac{1}{2\pi c} \left| \mathbf{k}_{n\omega_0} \times \int \mathbf{j}_n^{\rm col} \exp(-i\mathbf{k}_{n\omega_0} \cdot \mathbf{r}) d\mathbf{r} \right|^2 do. \quad (33)$$

In conclusion, we give an estimate of the total radiation intensity at the nth harmonic I_n by assuming that the collisional-current distribution (strong-field region) is uniform inside a cylindrical rod of radius a and length L (the z axis is directed along the pumping-field wave vector $\mathbf{k}_{\omega_0}$):

$$I_n = \frac{(j_n^{\rm col})^2}{2\pi c} (k_{n\omega_0})^2 \delta_n. \quad (34)$$

Here,

$$\delta_n = \left|\int \exp[i(n\mathbf{k}_{\omega_0} - \mathbf{k}_{n\omega}) \cdot \mathbf{r}]d\mathbf{r}\right|^2$$

is the factor responsible for the radiation direction, which is

$$\delta_n = \left|\frac{4\pi a^2}{k_\perp a} J_1(k_\perp a)\int_0^L \exp[i(nk_0 - k_z)z]dz\right|^2, \quad (35)$$

where $J_1(k_\perp a)$ is the Bessel function of the first kind.

It is easy to see that δ_n (and, hence, the intensity) is at a maximum if the following condition for the direction of the wave vector $\mathbf{k}_{n\omega_0}$ is satisfied [14]:

$$\frac{\mathbf{k}_{n\omega_0}\mathbf{k}_{\omega_0}}{n\mathbf{k}^2_{\omega_0}} = \frac{k_z}{nk_0} = 1. \quad (36)$$

In this case, the fields from all flashes (plasma collisions) will add up in amplitude rather than in intensity, as in the case of ordinary incoherent bremsstrahlung. For plasma, this condition is equivalent to the wave propagation at such an angle θ_0 to the propagation direction of the pumping wave that

$$\cos\theta_0 = \frac{\sqrt{1-\omega_p^2/\omega_0^2}}{\sqrt{1-\omega_p^2/(n\omega_0)^2}} = 1 - \left(\frac{1}{2} - \frac{1}{2n^2}\right)\frac{\omega_p^2}{\omega_0^2}, \quad \omega_0 \gg \omega_p. \quad (37)$$

Here,

$$\omega_p^2 = 2\pi e^2 n/m$$

is the electron plasma frequency, Thus, we have for $k_\perp$

$$k_\perp \approx \omega_p/c. \quad (38)$$

We immediately see from this relation that the dependence on $k_\perp a = \omega_p a/c$ in (35) is nonmonotonic; consequently, the density dependence in (34) is more complex. In particular, the total radiation $a \sim c/\omega_p$ intensity is proportional to the square of the density.

The cone thickness (Fig. 7) can be easily estimated from the condition

$$\delta\theta_\omega \approx \frac{1}{\sqrt{k_\omega a}}. \quad (39)$$

It should also be noted that, since the sizes a and L themselves (Fig. 7) depend on the field and plasma density, the resulting dependence on pumping intensity and density can significantly differ from the dependence derived by using the coherent bremsstrahlung cross section (31). In particular, allowance for the self-action effects can lead to a dependence on the field intensity in the form

$$I \propto E^4 \chi_{\mathrm{coh}}(E). \quad (40)$$

REFERENCES

1. I. P. Christov, M. M. Murnake, and H. C. Kapteyn, Phys. Rev. Lett. **78**, 1251 (1997); P. B. Corcum, M. H. Burnett, and M. Yu. Ivanov, Opt. Lett. **19**, 1870 (1994).
2. M. V. Fedorov, *Electron in a Strong Optical Field* (Nauka, Moscow, 1991).
3. V. P. Silin, Zh. Éksp. Teor. Fiz. **47**, 2254 (1964) [Sov. Phys. JETP **20**, 1510 (1964)].
4. V. P. Silin, Zh. Éksp. Teor. Fiz. **114**, 864 (1998) [JETP **87**, 468 (1998)].
5. F. V. Bunkin, A. E. Kazakov, and M. V. Fedorov, Usp. Fiz. Nauk **107**, 559 (1972) [Sov. Phys. Usp. **15**, 416 (1972)]; F. V. Bunkin and M. V. Fedorov, Zh. Éksp. Teor. Fiz. **49**, 1215 (1965) [Sov. Phys. JETP **22**, 844 (1965)].
6. G. Shvets and N. J. Fisch, Phys. Plasmas **4**, 428 (1997).
7. G. J. Pert, J. Phys. B **12**, 2756 (1979).
8. J. Dawson and C. Oberman, Phys. Fluids **5**, 517 (1962).
9. N. M. Kroll and K. M. Watson, Phys. Rev. A **8**, 804 (1973).
10. M. H. Mittleman, *Introduction to the Theory of Laser-Atom Interactions* (Plenum, New York, 1993).
11. G. M. Fraiman, V. A. Mironov, and A. A. Balakin, Phys. Rev. Lett. **82**, 319 (1999); G. M. Fraiman, V. A. Mironov, and A. A. Balakin, Zh. Éksp. Teor. Fiz. **115**, 463 (1999) [JETP **88**, 254 (1999)].
12. A. A. Balakin and G. M. Fraiman, Prikl. Fiz. **3**, 154 (2000); G. M. Fraiman and A. A. Balakin, Physica D (Amsterdam) **152–153**, 731 (2001).
13. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Nauka, Moscow, 1988; Pergamon, Oxford, 1975).
14. S.-Y. Chen, A. Maksimchuk, and D. Umstadter, Phys. Rev. Lett. **84**, 5528 (2000).
15. G. L. Yudin and M. Yu. Ivanov, Phys. Rev. A **63**, 033404 (2001).
16. S. Pfalzner and P. Gibbon, Phys. Rev. E **57**, 4698 (1998).

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 93, No. 4, 2001, pp. 706–716.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 810–822.

ATOMS, SPECTRA, RADIATION

The Primary Stages of the Charge Carrier Photogeneration in C_{60} Films Studied by the 100-fs Laser Pulse Pump–Probe Method

S. V. Chekalin[a, *], A. P. Yartsev[b], and V. Sundström[b]

[a]*Institute of Spectroscopy, Russian Academy of Sciences, Troitsk, Moscow oblast, 142190 Russia*
**e-mail: chekalin@isan.troitsk.ru*
[b]*Department of Chemical Physics, Lund University, S-22100 Lund, Sweden*
Received March 23, 2001

Abstract—The primary stages of photoinduced processes are studied in thin C_{60} films by the femtosecond laser pump–probe method. The films were excited by 100-fs laser pulses with photon energies above (wavelengths 345 and 367 nm) and below (645 nm) the mobility threshold, the fraction of excited molecules being no more than several percent. Upon probing in the spectral range from 400 to 1100 nm, several regions with substantially different decay kinetics were observed in the difference spectrum, which is caused by the simultaneous presence of several relaxing components. The appearance of the 465- and 500-nm bleaching bands in the difference spectrum upon excitation by photons with energies both above and below the mobility threshold, which are typical for electroabsorption spectra, suggests that charge carriers are produced in both these cases. The observed dependence of relaxation on the oxygen amount in the sample volume suggests that during excitation both charged (electrons and holes) and neutral (excited molecules) components are produced. The fraction of charged components is greater upon excitation into the fundamental band. The appearance of the 500-nm absorption band delayed by 10^{-13}–10^{-14} s, the delay being increased in the presence of oxygen, was attributed to the formation of excited anions due to the capture of electrons by C_{60} molecules. It is concluded that upon excitation of the films by photons with the energy below the mobility threshold, charge carriers are produced due to two-photon absorption rather than due to singlet–singlet annihilation. When the films are excited by photons above the mobility threshold, the primary charge carriers are produced by direct optical excitation. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The properties of fullerenes—molecules consisting of carbon atoms that form a closed spherical or spheroidal shell—have recently attracted special attention. The C_{60} molecule, which has the symmetry I_h, is the most spherical molecule among all the known molecules. The extensive studies of this molecule are related not only to its geometrical perfection but also to a number of other remarkable properties. The delocalization of π electrons inherent in the fullerene molecule due to a great number of conjugated bonds strongly enhances a nonlinear optical response of fullerenes. For this reason, the nonlinear optical susceptibility $\chi^{(3)}$ of fullerenes in solutions and solids has been studied in many papers [1–6]. Photoinduced darkening of fullerenes, which is observed in a broad spectral range, makes them a promising material for optical limiters [7, 8]. Another unique property of fullerenes is their ability to absorb many photons without decomposition of the molecule [8–10]. A very high triplet-state yield of fullerenes in solutions (the quantum yield of this process is close to unity) resulting in a high quantum yield of singlet oxygen makes C_{60} a potential agent for photodynamic therapy [11]. The discovery of superconductivity in C_{60} doped with alkali metals [12] stimulated the development of studies devoted to synthesis of organic superconductors. Thin polymer films doped with C_{60} exhibit photoconductivity with a very high quantum yield [13], which makes them promising for applications in xerography and in light cells.

An increase in the photoconductivity of a polymer upon doping with C_{60} is directly related to photoexcitation of C_{60} followed by the electron exchange with an adjacent polymer chain. It is known that pure C_{60} does not contain free charge carriers. Therefore, to induce conductivity, the C_{60} molecules should be excited, for example, by light. To reveal the features of charge transfer in fullerene-doped materials, it is necessary to study in detail the mechanisms of photoconductivity in pure fullerene films, and such studies were performed in many papers [14–19]. Note, however, that the results obtained so far are contradictory. For example, the quantum yield of charge carriers measured in different papers varied from 10^{-5} to 0.55, and the mobility of charge carriers estimated in different papers was also substantially different [18]. A great scatter in the measurements of these quantities is mainly explained by a high sensitivity of photoconductivity to the degree of

crystallinity of a sample, to the number of defects, and especially to the presence of oxygen, which can reduce photoconductivity by several orders of magnitude [15, 17, 18]. In addition, the phototransformation of fullerenes during measurements and the appearance of long-lived localized charge carriers, whose lifetime can achieve a week, can substantially affect the results of measurements [18, 19]. Finally, one of the important factors affecting the measurements is the use of metal electrodes and rather strong electric fields in most experiments, resulting in the unavoidable contact and electroabsorption effects. Based on the measurements of spectral dependences of the efficiency of generation of carriers, the authors of papers [14, 15] concluded that upon excitation above the mobility threshold (2.3 eV), charge carriers are generated due to direct optical excitation, while upon excitation at lower energies, they are generated due to singlet–singlet annihilation of excited molecules. However, the time resolution of these experiments was in the picosecond range, which is insufficient for a direct observation of primary photoprocesses in solids.

Relaxation of photoexcited C_{60} molecules in films has been studied by the pump–probe method with a femtosecond resolution in papers [1–3, 6, 20–34] (review of these experiments is presented in [34]). In most papers, a nonexponential decay of the photoinduced absorption was observed with the decay times of different components ranging from subpicoseconds to hundreds of nanoseconds. Many experiments [1, 3, 6, 20–23] were performed using the same wavelength for pumping and probing; in some studies [24–34], the probing was performed with the help of a broadband supercontinuum pulse, which gives much more information. The results of these experiments are sometimes substantially different or even contradictory in details. Thus, in some experiments, the dependences of relaxation on the pump intensity [1, 6, 21, 23, 25, 30–34], the pump wavelength [1], the probe wavelength [25, 31–34], and temperature [23] were observed, whereas in other experiments ([28, 29], [31], [26, 28, 30], [20, 21], respectively), such dependences were not observed. This discrepancy is probably mainly explained by different experimental conditions. First, in these experiments different samples were used. For example, spectral properties of C_{60} films, especially in the wavelength range above 600 nm, substantially depend on the conditions of their deposition [35]. Experimental results also depend on the amount of oxygen in films and on the dose of absorbed light [24]. Second, the results also may depend on the pulse repetition rate of a laser used in experiments. In the case of megahertz pulse repetition rates, which were used in papers [20, 22–24, 27], long-lived photoproducts can be accumulated during the experiment [36]. These can be triplet molecules [36], photopolymerized molecules [24], or localized charges [30]. The accumulation of photoproducts is not very important at pulse repetition rates less than a kilohertz and in samples containing oxygen, which efficiently quenches triplet molecules and inhibits photopolymerization [1, 37].

All these factors gave rise to the use of various mechanisms for explaining femtosecond relaxation of C_{60} in films observed in different experiments. This is the capture of carriers in the case of disordered semiconductors [23], hopping transfer of carriers [20], scattering of carriers and relaxation of a lattice [3, 21, 28, 33], distortion of a lattice upon exciton self-trapping in conjugated polymers [21], intramolecular relaxation [22], and the interaction and annihilation of excitons [1, 6, 25–27, 29–32, 34]. An important feature in the interpretation of all these experiments is the assumption that only one component appears upon excitation, which can transform to other intermediate and photoproducts during its relaxation. The nature of this primary component was treated differently in different papers (a Frenkel exciton [27, 29], a charge-transfer exciton [38], a molecule in an excited state [1, 3, 6, 21, 22, 24–26, 30–32], charge carriers [23, 28, 33], etc.), and various relaxation mechanisms were proposed. On the other hand, it follows from photoconductivity experiments that, because the quantum yield of the generation of charge carriers is always less than unity, direct optical excitation should produce both charged and neutral components, the types of their relaxation being substantially different. When relaxation is studied by the pump–probe method, different components (and their relaxation mechanisms) can be observed upon probing in different spectral regions. In this case, specific features of these mechanisms can be studied comparatively simply, even if the relaxation of different components occurs on the same time scale. However, a strong overlap of broad spectral lines of photoproducts produced upon excitation, which is typical for solids, can make the measurements very complicated. The single-component interpretation used in most papers is probably explained by the fact that the probing in these papers was performed only at a single wavelength, usually coinciding with the pump wavelength. In this case, the presence of many components in a system of excited molecules in principle cannot be found. In most (if not in all) experiments with broadband probing, the pump-beam intensity provided excitation of more than 10% of all molecules. At such a high fraction of excited molecules, the main channel of their relaxation is the singlet–singlet annihilation of neighboring molecules [1, 25, 30–32, 34], which substantially complicates the study of other fast decay channels.

In this paper, we describe pump–probe experiments with broadband probing. The fraction of excited molecules in our experiments did not exceed a few percent, which allowed us to decrease substantially the contribution of annihilation processes and to observe the dynamics of the appearance and decay of primary photoproducts, both neutral and charged, for different excitation levels. We also investigated the influence of oxygen on the relaxation and analyzed the conditions at which the effect of accumulation of photoproducts

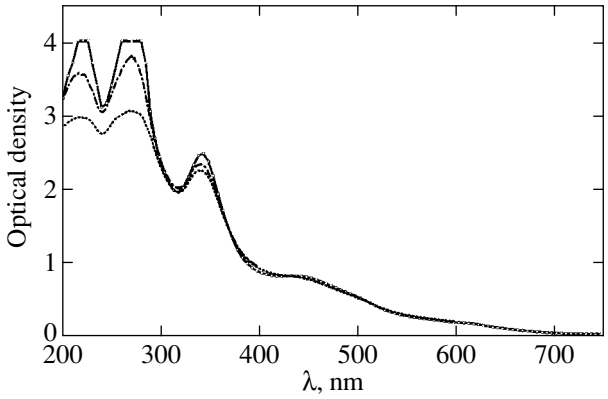

Fig. 1. Linear absorption spectrum of the C_{60} film of thickness 140 nm used in experiments. The solid curve is the spectrum immediately after the film deposition; the dot-and-dash and dotted curves are spectra after 30 and 50 h of irradiation by laser pulses during experiments, respectively.

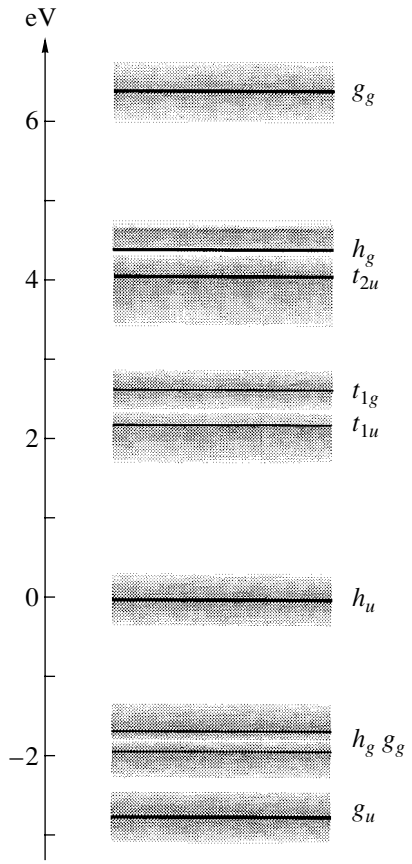

Fig. 2. Energy level diagram of the C_{60} film in the spectral region under study.

becomes substantial. We studied the relaxation beginning from the earliest stages up to 0.5-ns delays under the same pumping and probing conditions, which allowed us to identify the nature of the relaxing components.

2. EXPERIMENTAL

We studied the dynamics of difference absorption spectra of C_{60} films of thickness 140–190 nm deposited on quartz substrates by the pump–probe method. The films were excited by 100-fs pulses at 645 nm (10^{10}–10^{11} W/cm^2), 345 and 367 nm (10^9 W/cm^2). The energy absorbed by a sample was also measured.

A typical absorption spectrum of the C_{60} film shown in Fig. 1 is similar to the spectra described in earlier papers. The three intense absorption bands with maxima at 345, 270, and 220 nm (Fig. 1) are related to the $h_g g_g$–t_{1u}, h_u–h_g, and $h_g g_g$–t_{2u} transitions, respectively (Fig. 2), in accordance with notation accepted in previous papers. A group of weak lines near 600 nm is related to forbidden molecular h_u–t_{1u} (HOMO–LUMO) transitions, which are observed due to the Herzberg–Teller coupling. It was shown in [39] that the temperature dependences of the intensities of the 450- and 600-nm bands are substantially different. For this reason, the 450-nm band, which is related to the solid-state effects [40], was assigned to the allowed h_u–t_{1g} transition to the excited states with symmetry T_{1u}, T_{2u}, H_u, and G_u. The dipole nature of this allowed transition was also confirmed in recent experiments. The wavelength 645 nm corresponds to the h_u–t_{1u} excitation (below the mobility threshold), while the wavelength 367 nm (in some experiments, excitation was also performed at 345 nm) corresponds to the $h_g g_g$–t_{1u} excitation (above the mobility threshold).

The probing was performed by a weak femtosecond supercontinuum pulse in the spectral range from 400 to 1100 nm using delays from −1 to 550 ps. The pulse repetition rate was 5 kHz. The accuracy of measurements of difference spectra provided by the detection system was no worse than 10^{-4} (in optical density units), each experimental point being obtained by summation of 2500 pump pulses. The measurement errors were determined in each specific experiment by a degree of the sample homogeneity. All the experiments were performed at room temperature.

During the near-zero delay measurements, a signal detected from a sample also contains the substrate response [25], whose contribution can be substantial. For this reason, we measured a signal from a substrate in all experiments separately. Figure 3 demonstrates the relation between signals from a sample and a substrate upon excitation at 345 and 645 nm. One can see that in the latter case (Figs. 3a, 3b), the contribution from a substrate during the near-zero delay measurements is rather large (Figs. 3a, 3b). Nevertheless the accuracy of measurements allowed us to take this contribution into account and perform the correction of relaxation kinetics and difference spectra. All the experimental data presented below were corrected taking the above-mentioned contribution into account. In addition, by detecting a signal from the substrate, we could determine the

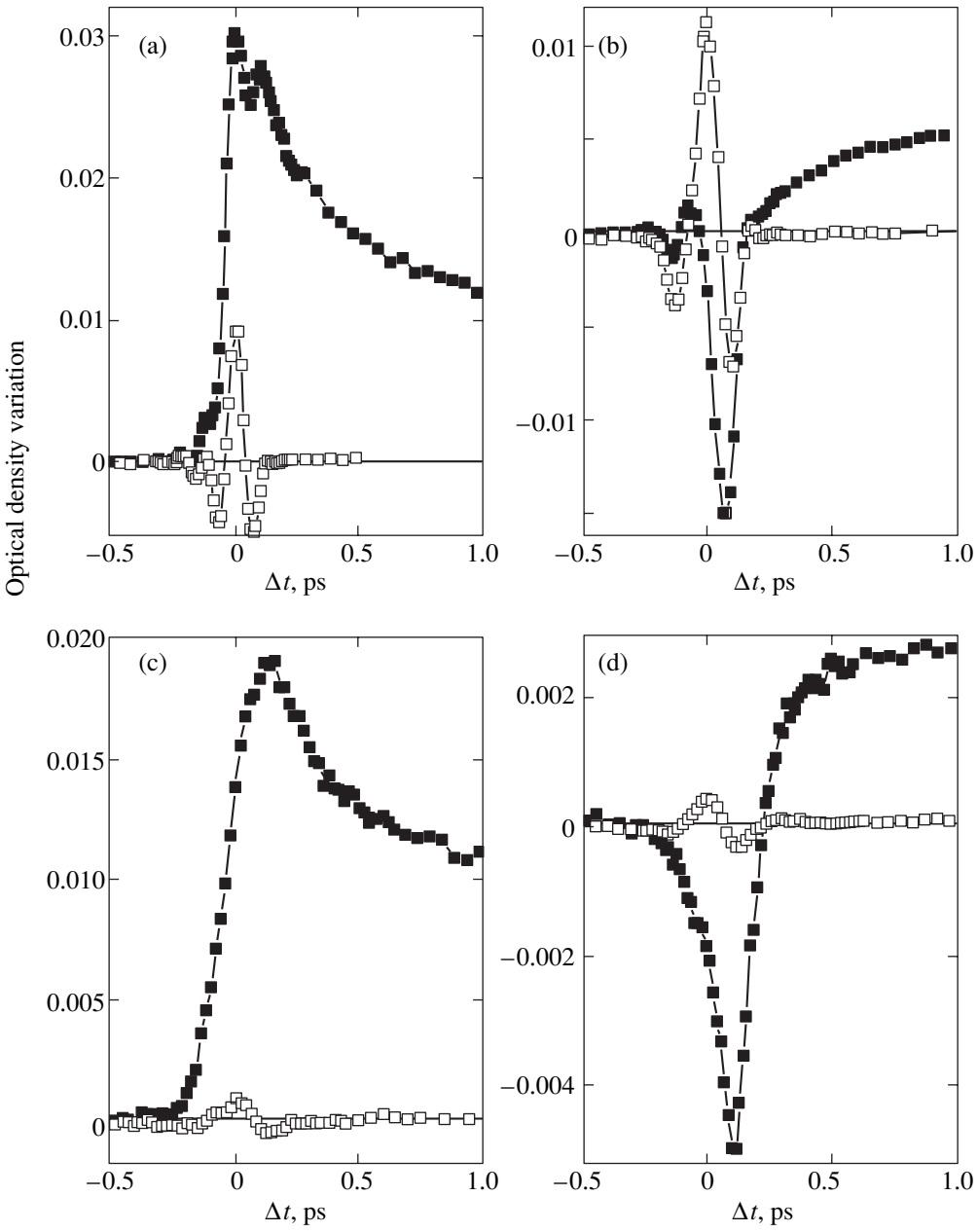

Fig. 3. Variation in the optical density of the difference spectrum of a sample (dark squares) and a substrate (light squares) near the zero delay. Excitation at 645 nm and probing at (a) 550 and (b) 500 nm; excitation at 345 nm and probing at (c) 550 and (d) 500 nm.

position of a zero delay in the relaxation kinetics upon probing at different wavelengths.

Because the pump pulse repetition rate was rather high (5 kHz), we performed special measurements of the effects of accumulation of photoproducts during our experiments [36]. It was found that upon excitation of a fixed sample at 367 and 345 nm, the relaxation kinetics changed substantially with increasing exposure (Fig. 4a), the rate of the change upon excitation of the films from the substrate side being several times larger than that upon excitation from the air side. This effect is obviously caused by the accumulation of photoproducts during excitation of the films. For this reason, the sample was continuously displaced in the subsequent experiments in the direction perpendicular to the laser beam. The rate of the sample movement was chosen so that the relaxation kinetics and difference spectra did not change with increasing exposure (Fig. 4b). In addition, a slower degradation of the sample was also observed upon laser excitation, resulting in a change in the linear absorption spectrum (Fig. 1). For this reason, linear spectra were regularly controlled and degraded samples were replaced by new samples.

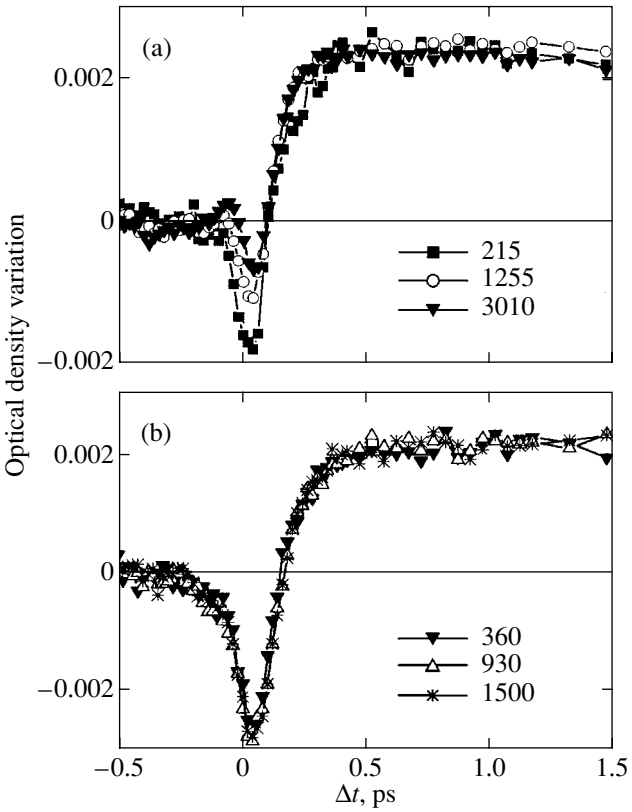

Fig. 4. Effect of the accumulation of photoproducts on the relaxation dynamics in the 500-nm band for the C_{60} film in the case of a fixed (a) and moving (b) sample. Excitation by 3×10^{-4} J cm^{-2} pulses was performed at 345 nm from the substrate side. The numbers of pump pulses (in thousands) incident on a sample during the measurements are indicated near the corresponding symbols.

Upon excitation into the fundamental absorption band at 345 and 367 nm, a great difference was observed in some spectral ranges of probing between the kinetics for films excited from the substrate side and from the air side (Fig. 5). In our opinion, this is explained by a different content of oxygen in the sample volume (see below). Samples were kept in the air; therefore, to reduce the influence of oxygen, we excited films from the substrate side.

3. EXPERIMENTAL RESULTS

We found that difference spectra exhibited the same main bands upon excitation at different wavelengths. Figure 6 shows the part of difference spectra, which is the most interesting from the point of view of the dynamics of charge carriers. The spectra were obtained upon excitation of the 190-nm thick film by laser pulses at 367 nm for several delays between the pump and probe pulses. Immediately after excitation, a decrease in the sample absorption (bleaching) is observed in the region from 430 to 520 nm, which has the complicated time dependence. Then, absorption increases within a broad band with a maximum at 550 nm. In a longer-wavelength region, a structureless absorption band appeared with a weakly pronounced maximum at 900 nm. The dynamics of appearance and relaxation of spectral changes was investigated in more detail from the relaxation kinetics obtained by probing samples at fixed wavelengths by continuously varying the delay. Analysis of the variation of the difference spectrum shows that several spectral regions can be distinguished that have substantially different relaxation kinetics (Fig. 7). This is clearly demonstrated at large time delays (Fig. 7a). The different types of relaxation observed in different spectral regions of probing (400–530, 450–700, and 700–1100 nm) under the same pumping conditions suggest that the nature of these characteristic bands is different (for simplicity, we denote these bands as the 500-, 600-, and 900-nm bands). This means that at least three different components simultaneously exist, each of them having its own relaxation kinetics. The characteristic bands of these components in the difference spectrum are strongly overlapped. In addition, the type of relaxation in different bands differently depends on the excitation intensity and on the amount of oxygen in the sample volume. Let us discuss the differences that were observed upon excitation at 367 (345) and 645 nm. First, this is the difference in the relaxation kinetics of the 500-nm band during the first several picoseconds (Fig. 8). Second, the ratio of maximum optical densities of the broad 900-nm band and of the most intense 600-nm band is approximately three times lower than upon excitation at 645 nm. In addition, in the latter case, the relaxation rate in the 600-nm band increased with increasing excitation intensity, but it was always lower than upon excitation at 367 nm, where such dependence was not observed. In both cases, the relaxation rate in the 900-nm band increased with increasing excitation intensity, whereas in the 500-nm band, it was virtually independent of the excitation intensity. As for excitation of samples from opposite sides, the different results were obtained only upon excitation at 367 and 345 nm. Upon excitation from the air side, the relaxation rate of the 500- and 900-nm bands decreased at the initial stage and then increased at delays of the order of tens of picoseconds (Figs. 5a, 5b). Relaxation in the 600-nm band was independent of the direction of the pump beam for any delay (Fig. 5c).

One of the most important properties of the difference spectra observed for all excitation wavelengths used in experiments is the appearance during the pump pulse of specific features in the spectral range from 400 to 530 nm (Fig. 6, dips in absorption at 465 and 500 nm), which are clearly observed at the maximum delay 550 ps as well (Fig. 6b). Note that these features and absorption at 600 and 900 nm appear with the rising front, which is determined by the pump pulse duration. Then, the shape of the difference spectrum in the 600-nm band changes: its short-wavelength wing shifts to the

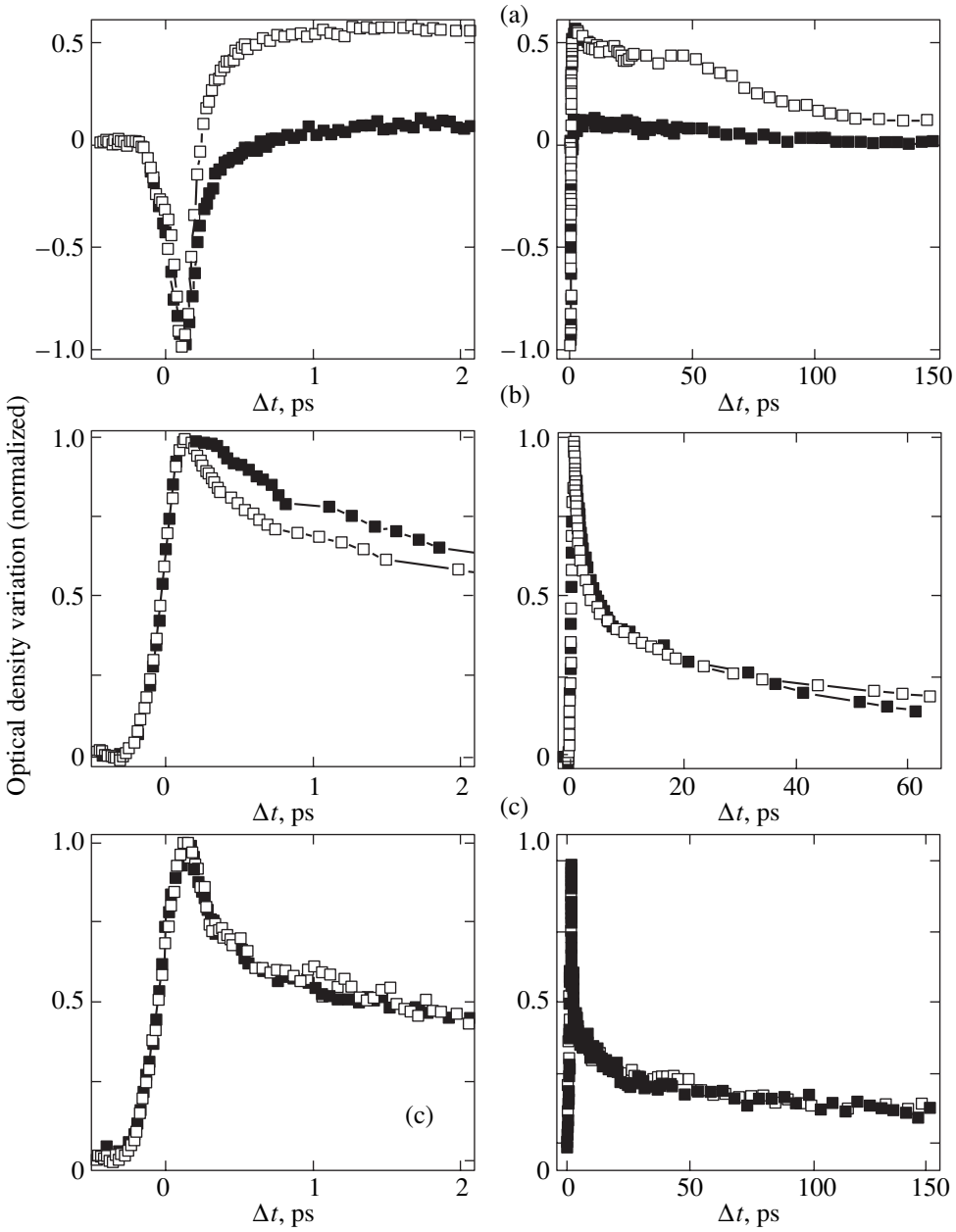

Fig. 5. Relaxation kinetics in (a) the 500-, (b) 900-, and (c) 600-nm bands obtained upon pulsed excitation at 367 nm from the substrate side (light squares) and from the air side (dark squares). All the dependences are normalized to the maximum.

blue, while the shape of the long-wavelength wing remains virtually unchanged (Fig. 9). Figure 10 shows the results of subtraction of the difference spectrum with the delay 0.1 ps from the spectra with larger delays (all difference spectra were preliminary normalized to the maximum absorption at 560 nm). These data clearly show that variations in the shape of the difference spectrum in Fig. 9, beginning from the delay 100 fs, are related to the appearance of the 500-nm absorption band. Figures 5a, 7b, and 8 show the detailed dynamics of this absorption and of photoinduced variations in the 500-nm band at earlier stages under different excitation conditions. In all the relaxation kinetics, the bleaching that appears at the initial moment rapidly changes to absorption. Upon excitation at 345 and 367 nm, the bleaching front is determined by the leading edge of the probe pulse. One can see from Fig. 5a that the front of the absorption kinetics in the 500-nm band is delayed, the delay being greater and induced absorption being smaller upon excitation from the air side. Upon excitation at 645 nm, the dynamics of bleaching was more complicated (Fig. 8), while the dynamics of absorption was independent of the side from which the sample was excited.

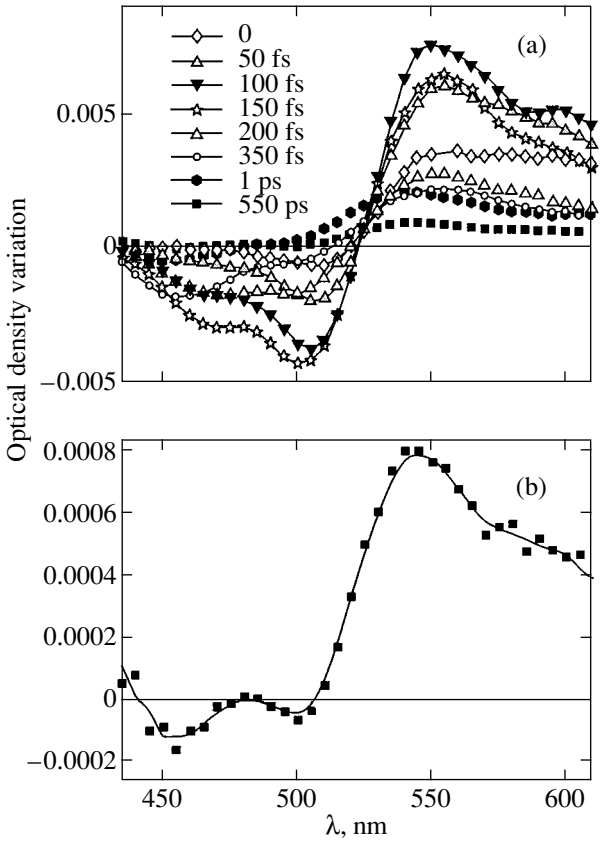

Fig. 6. (a) Difference spectra observed upon excitation of the C_{60} film of thickness 190 nm by 100-fs pulses at 367 nm (3×10^{-4} J cm^{-2}) for different delays. The delay values are shown in Fig. 6a. (b) The difference spectrum for the maximum delay 550 ps at the enlarged scale.

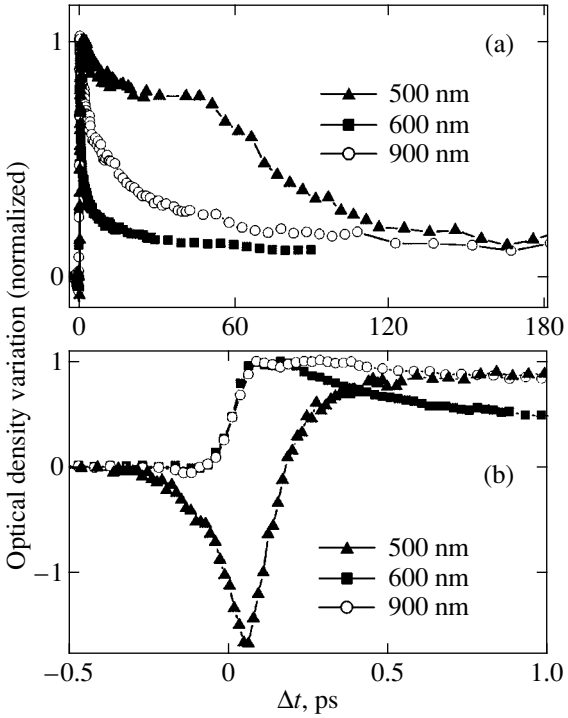

Fig. 7. Typical relaxation kinetics observed in the 500-, 600-, and 900-nm bands for (a) small and (b) large delays. All the dependences are normalized to the maximum absorption.

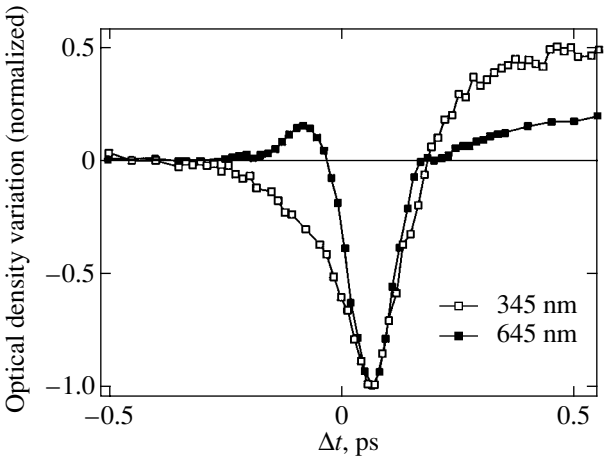

Fig. 8. Relaxation kinetics in the 500-nm band near the zero delay upon excitation at 645 and 345 nm. The dependences are normalized to the maximum.

We observed a linear dependence of the energy absorbed in films on the incident energy for all excitation wavelengths used in experiments.

4. DISCUSSION OF RESULTS

The characteristic bleaching with maxima at 465 and 500 nm in the difference spectrum of C_{60} films was earlier observed upon excitation above the mobility threshold at 530 and 478 nm [33] and 419 and 383 nm [38]. The same spectral features were observed in photomodulation experiments [41], where the similarity between this region of the difference spectrum and the electroabsorption spectrum of C_{60} films was noted [42]. Based on this similarity, the spectral features were attributed to electroabsorption changes in the spectra of molecules in the ground state in local electric fields produced by charge carriers [43]. These changes ΔD are proportional to the square of the electric field and, in the case of an isotropic medium with randomly oriented molecules, have the form

$$\Delta D = \frac{1}{2}\frac{\partial D}{\partial E}\Delta p F^2 + \frac{1}{6}\frac{\partial^2 D}{\partial E^2}(\Delta m)^2 F^2, \qquad (1)$$

where Δp and Δm are changes in the polarizability and dipole moment of the molecule in the electric field F and $D(E)$ is the dependence of the optical density on the photon energy (linear absorption spectrum).

The bleaching at 465 and 500 nm, which was observed in experiments [38] performed at 80 K, was attributed to excitation of an exciton accompanied by the charge transfer, and the following relaxation was interpreted as the exciton self-trapping and its transition to a long-lived triplet state. It seems more likely that the bleaching bands observed in our experiments have the electroabsorption nature, because bleaching of the 450-nm band (Fig. 1) caused by the depletion of the ground state upon excitation would produce a smooth band with a maximum at 450 nm in the difference spectrum rather than two bands observed in experiments (Fig. 6). In addition, recent study [44] showed that only the 500-nm band belongs to a charge-transfer exciton, whereas the 465-nm band is related to the dipole transition of a Frenkel exciton. In contrast to [38], we did not observe a long-lived (lifetime about of 250 ns) self-trapped exciton; in any case, the relaxation of the 500-nm band mainly occurs at the 100-ps scale (Fig. 7a).

The electric field of a pump laser pulse also can produce electroabsorption changes, which are proportional to the square of the field [see (1)], however, this contribution is negligible under our experimental conditions. This was proved in special experiments, in which no signal was observed at 500 nm upon excitation of the same—C_{60} films at 800 and 1280 nm. Based on the above discussion, we can conclude that the appearance of the features in the difference spectrum, which are typical for the electroabsorption spectrum of C_{60} films [42] and exist up to maximum delays (Fig. 6), suggests that local electric fields are produced in a sample due to generation of charge carriers upon optical excitation at 367, 345, and 645 nm. One can see from the difference spectra in Fig. 6 that the dynamics of electroabsorption changes near the zero delay is rather complex: bleaching at 500 nm occurs more rapidly than at 465 nm and reaches larger maximum values (for the 150-fs delay). Then, this bleaching also more rapidly decreases. Such a behavior of the difference spectrum can be related to the changes in polarizability and the dipole moment of molecules in the ground state produced by the electric field. Another possible explanation is the appearance of a short-lived charge-transfer exciton, which results in the additional bleaching of the 500-nm band. In any case, to interpret physical processes proceeding at this time scale, experiments with a better time resolution are required.

Conduction electrons are majority charge carriers in C_{60} films [17, 18]. It is reasonable to expect that upon excitation of C_{60} molecules, electrons should appear together with cations (holes). The electric field of the charges produced causes electroabsorption changes in the spectrum of unexcited molecules (Fig. 6) due to changes in their polarizability and dipole moment (see formula (1) [43]). In addition, neutral molecules in the excited state should appear upon excitation because the quantum yield of charge carriers is less than unity.

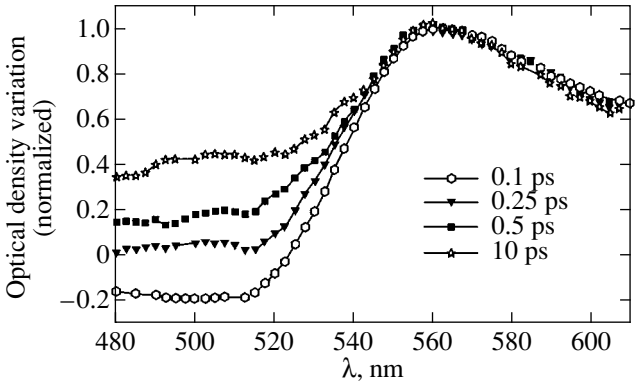

Fig. 9. Shapes of difference spectra for different delays. All the spectra are normalized to the maximum.

Unfortunately, the data on the spectra of cations and anions in C_{60} films are in fact absent in the literature. In solutions, both ions absorb in the entire visible and near-IR regions and have similar spectra with maxima in the near-IR region. The absorption cross section for a cation is smaller and its characteristic bands lie in a shorter-wavelength spectral region (between 800 and 1000 nm compared to the region from 950 to 1100 nm for an anion) [45]. It is reasonable to assume that due to the solid-state effects the characteristic absorption bands of charge carriers in films will be shifted to the blue compared to those in solutions, as in the case of neutral molecules [39]. However, because the magnitude of these shifts is unknown, we identified the bands in the difference spectrum using the fact that charged and neutral components differently interact with oxygen. It is well known that oxygen drastically affects the photoconductivity because it produces in C_{60} very efficient deep traps for charge carriers [15, 17, 18]. Because oxygen is accumulated under normal conditions in freshly prepared C_{60} films at the depth of the order of 20 nm from the film–air interface [46], its influence can be observed in experiments with short-wavelength excitation at which the absorption depth is substantially less than the film thickness (for excitation at 367 nm, the absorption depth is about 40 nm). Therefore, by comparing the data obtained upon excitation of the film at 367 and 345 nm from the air and substrate sides, we can determine the influence of oxygen on the dynamics of processes under study. These studies revealed the effect of oxygen on the relaxation kinetics of the 500-nm band (Fig. 5a) and, in somewhat lesser degree, on the relaxation kinetics of the 900-nm band (Fig. 5b). However, the relaxation kinetics of the 600-nm band was virtually independent of the presence of oxygen (Fig. 5c). This means that the first two kinetics reflect the relaxation of charged components, while the latter kinetics is related to neutral components, i.e., excited C_{60}* molecules. The 600- and 900-nm bands and electroabsorption changes in the spectrum

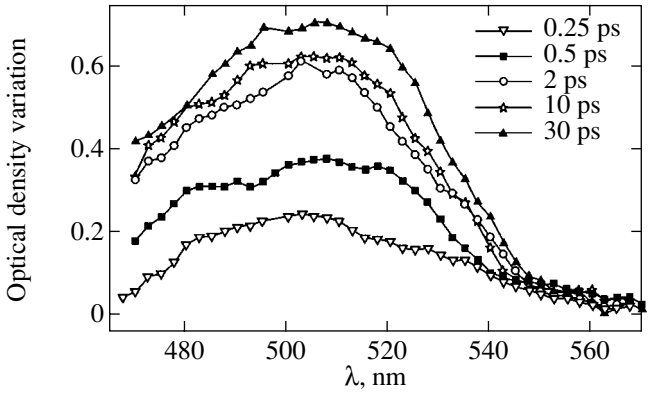

Fig. 10. Dynamics of the appearance of the 500-nm absorption band in the difference spectrum for delays longer than 100 fs (the corresponding delays are shown in figure).

appear with the rising front, which is determined by the pump pulse duration (Fig. 7b). For this reason, we assigned the 900-nm band to the C_{60}^{+} cation, which appears simultaneously with a conduction electron.

The rapid decay of bleaching at 500 nm (Figs. 5a, 7b, 8) was interpreted in [33] as a decrease in the influence of local fields (and, hence, the disappearance of electroabsorption changes produced by them) due to their screening with increasing carrier density during the pump pulse. We observed in our experiments the increase in bleaching with increasing excitation intensity, which contradicts to the screening mechanism. In addition, as was mentioned above, electroabsorption measurements were observed in the difference spectra up to the largest delays (Fig. 6). Therefore, we explain this part of the kinetics not by the bleaching relaxation but by the appearance of additional absorption at 500 nm with a strongly delayed front (see above). One can see from Fig. 5a that the duration of this front strongly depends on the presence of oxygen, which itself has a low mobility. This means that a rapid increase in absorption is related to a charge carrier having a high mobility, i.e., a conduction electron e^{-}, whose mobility is several orders of magnitude higher that of a C_{60}^{+} hole [17, 18]. We explain the increase in absorption at 500 nm, which was observed for delays from 0.1 to 10 ps (Fig. 10), by the capture of e^{-} by unexcited molecules with the formation of an excited C_{60}^{-} anion, which rather strongly absorbs in this band. The high efficiency of this process is caused by the high electron affinity of C_{60} (2.65 eV [18]). In the presence of oxygen, a conduction electron is efficiently captured by deep traps, resulting in a decrease in the rate of formation and yield of anions (Fig. 5a). Simultaneously with the formation of excited anions upon the electron trapping, anions are relaxing and a much slower recombination of anions and cations occurs. This results in a decrease in absorption in the relaxation kinetics at 500 nm. The delay in the absorption decay observed in experiments (Fig. 7a) can be explained by a gradual decrease in the rate of formation of excited anions caused by a decrease in the concentration of conduction electrons.

The differences observed upon excitation at different wavelengths are probably related to different mechanisms of generation of charge carriers. Upon excitation at 367 and 345 nm, i.e., above the mobility threshold, carriers appear during the entire pump pulse and the electroabsorption bleaching exceeds the absorption of excited molecules at the t_{1u}–h_g transition. To generate free carriers upon excitation at 645 nm, the C_{60} molecule should absorb more than one photon. Assuming that the two-photon absorption coefficient at 645 nm is close to the values measured at 620 nm [5], we can conclude that in the energy range where the kinetics and difference spectra have been measured, the two-photon absorption cross section is of the same order of magnitude or somewhat greater than the linear absorption cross section measured with a spectrophotometer. Therefore, upon excitation at 645 nm, charge carriers are generated due to two-photon absorption mainly in the most intense part of the pump pulse. This is confirmed by a more complicated form of the relaxation kinetics at 500 nm in this case (Fig. 8). Two-photon absorption at the leading edge of the pump pulse is lower than one-photon absorption, so that the number of charge carriers is small, the main part of excited molecules being in the t_{1u} state, from which absorption at 500 nm is larger than from the ground state (the allowed t_{1u}–h_g transition). For this reason, absorption is observed in the difference spectrum. As the excitation intensity approaches the maximum of the pump pulse, two-photon absorption becomes dominant and charge carriers appear that cause the electroabsorption bleaching of the 500-nm band. Upon excitation at 367 and 345 nm, charge carriers and bleaching produced by them appear already at the leading edge. It is obvious that the quantum yield of generation of carriers at 645 nm should be substantially lower. This is confirmed by a lower ratio of the intensities of the 900- and 600-nm bands observed in our experiment compared to excitation at 345 and 467 nm and is consistent with the photoconductivity data [14–17]. A dominant role of two-photon absorption upon excitation at 645 nm under conditions of our experiment is also confirmed by the fact that the difference spectrum did not exhibit the intense 680-nm band, which was assigned to the $h_g g_g$–h_u absorption, which appears due to the production of vacancies at the h_u level caused by excitation from this level [33]. This means that a greater part of molecules are excited not upon a transition from the h_u level but upon the two-photon $h_g g_g$–t_{1g} transition. The kinetic energy of conduction electrons produced in this case are higher than upon excitation at 367 and 345 nm (the $h_g g_g$–t_{1u} transition), resulting in a more rapid formation of anions upon excitation at 645 nm (cf. the relaxation kinetics at 500 nm, Fig. 8), which was

observed in our experiments. The relaxation of uncharged excited molecules (the 600-nm band) can be related to the quenching of excitons by charge carriers due to the dipole–dipole interaction [47] or accompanied by emission of phonons [28].

Based on the above discussion, we can interpret the features of accumulation of photoproducts, which were observed upon excitation at 367 nm (Fig. 4a). We explain them by photopolymerization of C_{60} in films during a sufficiently prolonged irradiation [37]. This assumption is confirmed by the fact that the accumulation process strongly slows down in the presence of oxygen, which efficiently quenches triplet C_{60} molecules involved in photopolymerization [37]. In addition, the radiation dose, at which variations in the relaxation kinetics of fixed samples are observed, corresponds to absorption of more than 10^4 photons per molecule, in good agreement with photopolymerization data [36, 37, 46]. The increase in the rate of formation of excited anions observed during accumulation of photoproducts (and the corresponding decrease in the bleaching amplitude in the relaxation kinetics at 500 nm, Fig. 4a) can be explained by an increase in the mobility of free electrons due to an increase in the photoconductivity of C_{60} upon photopolymerization [48].

5. CONCLUSION

Our studies showed that upon excitation of C_{60} films by 100-fs pulses at 645 nm (10^{10}–10^{11} W/cm^2) and at 367 (345) nm (10^9 W/cm^2), both primary charge carriers - conduction electrons and localized cations producing local electric fields in the sample, and excited neutral molecules are formed during the pump pulse. The fastest relaxation processes involve the most mobile components—conduction electrons. These are the capture of electrons by C_{60} molecules accompanied by the formation of excited anions, recombination of electrons and cations, and the capture of electrons by deep traps (in the presence of oxygen). The recombination of localized cations and anions and their capture by deep traps occur much slower. Upon excitation at 367 and 345 nm, primary charge carriers appear due to direct optical excitation, whereas upon excitation at 645 nm, they are produced due to two-photon absorption, so that the quantum yield in the latter case is substantially lower. The singlet–singlet annihilation was not observed in both cases; therefore, annihilation processes do not play an important role in the generation of charge carriers.

ACKNOWLEDGMENTS

The authors thank G. Trinkunas for the useful discussion and N. F. Starodubtsev for his help in preparing samples. This work was supported by the Russian Foundation for Basic Research (project no. 00-02-16660), the Swedish Institute, and the Swedish Royal Academy of Sciences.

REFERENCES

1. S. R. Flom, R. Pong, F. J. Bartoli, and Z. H. Kafafi, Phys. Rev. B **46**, 15598 (1992).
2. G. B. Talaparta, N. Manickam, M. Samoc, *et al.*, J. Phys. Chem. **96**, 5206 (1992).
3. M. J. Rosker, H. O. Marcy, T. Y. Chang, *et al.*, Chem. Phys. Lett. **196**, 427 (1992).
4. W. Ji, S. H. Tang, G. Q. Xu, *et al.*, J. Appl. Phys. **74**, 3669 (1993).
5. I. V. Bezel, S. V. Chekalin, Yu. A. Matveets, *et al.*, Chem. Phys. Lett. **218**, 475 (1994).
6. S. R. Flom, F. J. Bartoli, H. Sarkas, *et al.*, Phys. Rev. B **51**, 11376 (1995).
7. L. Tutt and A. Kost, Nature **356**, 225 (1992).
8. J. Wray, K. Liu, C. Chen, *et al.*, Appl. Phys. Lett. **64**, 2785 (1994).
9. P. Wurz, K. Lykke, *et al.*, J. Phys. Chem. **96**, 10129 (1992).
10. D. Palit, A. Sapre, J. Mittal, and C. Rao, Chem. Phys. Lett. **195**, 1 (1992).
11. J. Arbogast, A. Darmanyan, C. Foote, *et al.*, J. Phys. Chem. **95**, 11 (1991).
12. A. Hebard, M. Rosseinsky, R. Haddon, *et al.*, Nature **350**, 600 (1991).
13. K. A. Wang, Nature **356**, 585 (1992).
14. C. H. Lee, G. Yu, D. Moses, *et al.*, Phys. Rev. B **48**, 8506 (1993).
15. D. Moses, C. H. Lee, B. Kraabel, *et al.*, Synth. Met. **70**, 1419 (1995).
16. S. Kazaoui, R. Ross, and N. Minami, Phys. Rev. B **52**, R11665 (1995).
17. S. Priebe, B. Pietzak, and R. Konnenkamp, Appl. Phys. Lett. **71**, 2160 (1997).
18. M. S. Dresselhaus, G. Dresselhaus, and P. C. Eklund, in *Science of Fullerenes and Carbon Nanotubes* (Academic, New York, 1996).
19. E. A. Katz, V. Lyubin, D. Faiman, *et al.*, Solid State Commun. **100**, 781 (1996).
20. R. A. Cheville and N. J. Halas, Phys. Rev. B **45**, 4548 (1992).
21. S. D. Brorson, M. K. Kelly, U. Wenschuh, *et al.*, Phys. Rev. B **46**, 7329 (1992).
22. S. B. Fleischer, E. P. Ippen, G. Dresselhaus, *et al.*, Appl. Phys. Lett. **62**, 3241 (1993).
23. T. Juhasz, X. H. Hu, C. Suarez, *et al.*, Phys. Rev. B **48**, 4929 (1993).
24. S. B. Fleisher, B. Pevzner, D. J. Dougherty, *et al.*, Appl. Phys. Lett. **69**, 296 (1996).
25. S. V. Chekalin, E. Akesson, V. Sundström, and V. M. Farztdinov, Pis'ma Zh. Éksp. Teor. Fiz. **58**, 296 (1993) [JETP Lett. **58**, 295 (1993)].
26. T. N. Thomas, R. A. Taylor, J. F. Ryan, *et al.*, Europhys. Lett. **25**, 403 (1994).
27. D. Dick, X. Wei, S. Jeglinski, *et al.*, Phys. Rev. Lett. **73**, 2760 (1994).

28. V. M. Farztdinov, Yu. E. Lozovik, Yu. A. Matveets, *et al.*, J. Phys. Chem. **98**, 3290 (1994).
29. T. W. Ebbesen, Y. Mochizuki, K. Tanigaki, and H. Hiura, Europhys. Lett. **25**, 503 (1994).
30. S. L. Dexheimer, W. A. Vareka, D. Mittlemen, *et al.*, Chem. Phys. Lett. **235**, 552 (1995).
31. S. V. Chekalin, in *Fast Elementary Processes in Chemical and Biological System*, Ed. by A. Tramer (American Inst. of Physics, Woodbury, 1996), AIP Conf. Proc. **364**, 162 (1996).
32. S. V. Chekalin, in *Femtochemistry. Ultrafast Chemical and Physical Processes in Molecular Systems*, Ed. by M. Chergui (World Scientific, Singapore, 1996), p. 649.
33. V. M. Farztdinov, A. L. Dobryakov, N. R. Ernsting, *et al.*, Phys. Rev. B **56**, 4176 (1997).
34. D. Boucher, S. V. Chekalin, S. A. Kovalenko, *et al.*, Proc. SPIE **3239**, 302 (1997).
35. V. Capozzi, G. Casamassima, G. F. Lorusso, *et al.*, Solid State Commun. **98**, 853 (1996).
36. S. V. Chekalin, Appl. Phys. Lett. **71**, 1276 (1997).
37. P. Zhou, A. M. Rao, K. A. Wang, *et al.*, Appl. Phys. Lett. **60**, 2871 (1992).
38. M. Ichida, A. Nakamura, H. Shinohara, *et al.*, Chem. Phys. Lett. **289**, 579 (1998).
39. J. Hora, P. Panek, K. Navratil, *et al.*, Phys. Rev. B **54**, 5106 (1996).
40. K. Sinha, J. Menendez, R. C. Hanson, *et al.*, Chem. Phys. Lett. **186**, 287 (1991).
41. X. Wei, Z. V. Vardeny, D. Moses, *et al.*, Synth. Met. **49–50**, 549 (1992).
42. B. C. Hess, D. V. Bowersox, S. H. Mardirosian, *et al.*, Chem. Phys. Lett. **248**, 141 (1996).
43. L. Sebastian, G. Weiser, and H. Bassler, Chem. Phys. **61**, 125 (1981).
44. M. Knupfer and J. Fink, Phys. Rev. B **60**, 10731 (1999).
45. C. A. Reed and R. D. Bolskar, Chem. Rev. **100**, 1075 (2000).
46. C. C. Eloi, J. D. Robertson, A. M. Rao, *et al.*, J. Mater. Res. **8**, 3085 (1993).
47. M. Pope and C. E. Swenberg, *Electronic Processes in Organic Crystals* (Clarendon Press, Oxford, 1982).
48. E. I. Terukov, V. Yu. Davydov, and O. I. Kon'kov, Pis'ma Zh. Tekh. Fiz. **22** (5), 71 (1996) [Tech. Phys. Lett. **22**, 213 (1996)].

Translated by M. Sapozhnikov

Journal of Experimental and Theoretical Physics, Vol. 93, No. 4, 2001, pp. 717–722.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 823–829.

ATOMS, SPECTRA, RADIATION

A Constant Phase Autocorrelation Interferometer: A New Method for Measuring the Homogeneous Emission Linewidth under Conditions of Prevailing Inhomogeneous Broadening

M. V. Lebedev

Institute of Solid-State Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia
e-mail: lebedev@issp.ac.ru

Received April 27, 2001

Abstract—We propose a new interferometric method of measuring the homogeneous emission linewidth for an ensemble of radiating oscillators. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The concepts of homogeneous and inhomogeneous emission linewidth are widely used in spectroscopy to describe the emission line shape for an ensemble of independent oscillators. The best-known example of such a system is a low-pressure gas whose atoms, which are excited, for instance, by electric discharge, radiate a set of spectral lines of the corresponding chemical element. The natural emission linewidth is generally very small; the inhomogeneous broadening attributable to the Doppler frequency shift due to the thermal motion of radiating atoms dominates. Since the homogeneous linewidth is directly related to the optical-transition probability, it is a crucial parameter of the radiating system. Methods of nonstationary, nonlinear optics, such as a light echo and four-wave mixing [1], are used in spectroscopy for its determination.

Here, we propose a method for measuring the homogeneous linewidth that does not use any nonlinear optical effects. The method is based on the application of a new scheme of a grating interferometer. Diffraction gratings are successfully used in interferometry for beam splitting and for an additional light monochromatization [2]. In the proposed scheme, the grating allows the wavelength independence of the phase difference between the interfering beams to be achieved. An optical scheme of the interferometer is shown in Fig. 1. A parallel beam of light falls on a diffraction grating *G*. The diffracted light is reflected from a flat mirror *M* mounted parallel to the grating and is again directed to the grating. It is easy to show that after a second diffraction by the grating, the deflected beam will propagate exactly parallel to the zero-order beam that underwent a mirror reflection from the grating. Consequently, an interference of the two beams can be observed in the far field. The path difference S_{10} between the beams is given by

$$S_{10} = \frac{2z}{\cos\phi}(1 + \sin\phi\sin\theta). \quad (1)$$

The diffraction angle ϕ is related to the wavelength λ of the incident light and to the grating spacing h by

$$\sin\theta + \sin\phi = \frac{\lambda}{h}. \quad (2)$$

Differentiating (1) and using (2), it is easy to obtain

$$\frac{dS_{10}}{d\lambda} = \frac{2z\lambda}{h^2\cos^3\phi} > 0. \quad (3)$$

Let us determine the change in the phase difference

$$\Phi = 2\pi\frac{S_{10}}{\lambda} = 2\pi N$$

between the interfering beams, where N is the order of interference, for a small change in the wavelength of the incident light:

$$d\Phi = 2\pi\left(-\frac{S_{10}}{\lambda^2} + \frac{1}{\lambda}\frac{dS_{10}}{d\lambda}\right)d\lambda. \quad (4)$$

The first term in this formula corresponds to the change in wave phase at the end of a fixed-length optical path, as is the case, for example, in a Michelson interferometer, while the second term results from the grating action. Note that the "Michelson" and "grating" contributions have opposite signs. If we require that $d\Phi = 0$ (phase-constancy condition), then we can easily

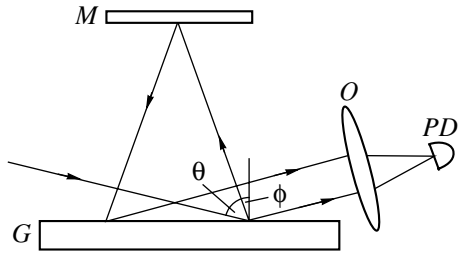

Fig. 1. An optical scheme of the constant phase autocorrelation interferometer: *G* is a grating, *M* is a flat mirror, *O* is an objective, and *PD* is a photodiode.

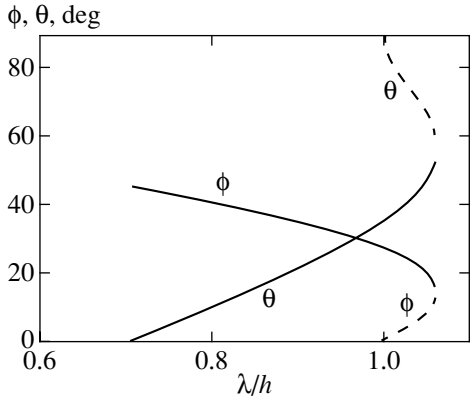

Fig. 2. The angles of incidence (θ) and diffraction (ϕ) for which the phase-constancy condition is satisfied; *h* is the grating spacing.

derive a fourth-degree equation for $x = \sin\phi$ using Eq. (2) for diffraction by a grating:

$$x^4 - \frac{\lambda}{h}x^3 - 2x^2 + \frac{\lambda}{h}x + \left(1 - \frac{\lambda^2}{h^2}\right) = 0. \qquad (5)$$

A numerical analysis of this equation with a computer shows it to have physical solutions in some range of λ/h The results of our computations are presented in Fig. 2, where the angles ϕ and θ that satisfy Eq. (5) are plotted against λ/h. Thus, the condition $d\Phi = 0$ can actually be met.

This condition implies that the phase difference between the interfering beams will not depend on a small change in wavelength, and, consequently, the frequency shift attributable to inhomogeneous line broadening will not cause the interference pattern to blur. Indeed, interference is known to be a single-photon effect; i.e., passing through the instrument, the photon emitted by a particular oscillator from an ensemble interferes with itself. The wave packet that corresponds to this photon has a carrier frequency characteristic of its emitting oscillator. If we directed this wave packet to a Michelson interferometer, then the probability to detect the photon at a certain point of the screen after interference would depend on the phase difference ΔΦ for the two paths along which this point is reached, with each corresponding to its own interferometer arm:

$$\Delta\Phi = 2\pi\frac{S_{12}}{\lambda}. \qquad (6)$$

Here, S_{12} is the difference between the optical interferometer arms, and λ is the carrier wavelength of the wave packet. For a wave packet with a different carrier wavelength λ', the phase difference ΔΦ' may turn out to be so large that the probability at the screen point where it was at a maximum for the first wave packet will be at a minimum for the second packet. Averaging over the wave packets radiated by all oscillators will cause the interference pattern to disappear.

In our interferometer, the path difference S_{10} depends on the carrier wavelength of the wave packet. As a result, the phase shift ΔΦ is the same for wave packets with different carriers, because increasing the wavelength causes a proportional increase in the path difference. In other words, an ensemble of quasi-monochromatic oscillators with slightly differing central frequencies will produce an interference pattern even if the inhomogeneous broadening is significant.

As we see from Eq. (5), the separation *z* between the mirror and the grating does not enter into the phase-constancy condition. Since a single-photon wave packet is spatially bounded, increasing this separation causes the interference to disappear. Thus, by changing *z* and observing the disappearance of interference, we can determine the spatial extent of a single-photon wave packet and, consequently, the radiative lifetime of the observed transition or the homogeneous linewidth.

The above considerations properly illustrate the main idea, but they are very simplified in two important respects. They disregard the delay introduced by the grating into the diffracted beam and the finite grating resolution. Let us consider each of these effects sequentially.

A diffraction grating is known to produce a diffracted beam with a delay that depends on the total number of grating lines and on the angles of incidence and diffraction. This effect is widely used in lasers to compress (shorten) laser pulses. A pulse-compression system was proposed by Treacy [3], who also considered a theory of this phenomenon. The system consists of two identical parallel diffraction gratings; basically, it is very similar to our interferometer (see Fig. 3a). Its main distinction is that it is not an interferometer proper; i.e., there is no addition of the transmitted signal with a reference beam that split off from the main one and that arrived at the same point by a different path. Otherwise, the beam path in Treacy's system is the same as that in our interferometer. This becomes clear if we construct a virtual image of the grating produced by the mirror (Fig. 3b). As a result, Treacy's reasoning proves to be also valid in our case if twice the separation 2*z* between the mirror and the grating is sub-

stituted for the separation between the gratings in his scheme.

In his calculations, Treacy disregarded the finite grating resolution. He showed that the pulse shape and duration at the output changed because of the difference in the diffracted-signal formation delays for different Fourier components of the pulse incident on the gratings. The pulse-compression problem turned out to be formally similar to the Fresnel diffraction problem if we ignore the term $d^2\tau/d\omega^2$ compared to $d\tau/d\omega$ in the range of frequencies ω concerned, where τ is the delay attributable to the wave-packet passage through the system. Treacy's estimates [3] indicate that for a pulse of spectral width $\delta\omega$, the ratio of the contributions from these two terms to the phase is $\delta\omega/\omega$ in order of magnitude; i.e., the corrections are actually negligible. Note that the above phase-constancy condition is written in the same approximation.

If a classical damped oscillator with a low damping, which is commonly used for a classical description of atomic radiation, is taken as the basis, then the wave packet can be represented as a weakly damped harmonic wave with a more or less steep leading edge that corresponds to oscillator "switching." Following Treacy's notation [3], we write the wave packet before it falls on the grating as

$$B_{\text{in}}(t) = A(t)\exp[i\psi(t)]\exp[-i\omega_0 t], \tag{7}$$

where $A(t)$ is a slowly varying amplitude, and $\psi(t)$ is the function that describes the pulse phase modulation. Equation (7) describes the signal time variation in an $x = 0$ plane parallel to the pulse wave front, which is assumed to be plane. After the passage through the two-grating system (or an equivalent system composed of a grating and a mirror in our interferometer), the pulse wave front turns out to be parallel to the same plane, while the time dependence is

$$B_{\text{out}}(t'+\tau_0) = \left(\frac{\mu}{2\pi}\right)^{1/2}\exp\left[i\left(\phi_0-\frac{\pi}{4}-\omega_0(t'+\tau_0)\right)\right] \times\int_{-\infty}^{\infty}A\exp(i\psi)\exp\left[i\frac{\mu}{2}(t-t')^2\right]dt. \tag{8}$$

Here, τ_0 is the delay at the carrier frequency ω_0, which is attributable to the wave-packet passage through the system, $\mu^{-1} = -\partial\tau/\partial\omega$, and ϕ_0 is a constant.

For a carrier frequency with a constant amplitude,

$$B_{\text{in}}(t) = A_0\exp(-i\omega_0 t) \tag{9}$$

we obtain at the output

$$B_{\text{out}}(t'+\tau_0) = A_0\exp[i(\phi_0-\omega_0(t'+\tau_0))]. \tag{10}$$

Note that the phase Φ introduced above corresponds to the carrier-wave phase $\omega_0\tau_0$ and differs from Treacy's phase [3] by the absence of a correction term that allows for the relative delay of various Fourier components (including this term is necessary when considering the pulse passage through the system). Thus, the phase-constancy condition $d\Phi = 0$ formulated above implies the constancy of the carrier-wave phase.

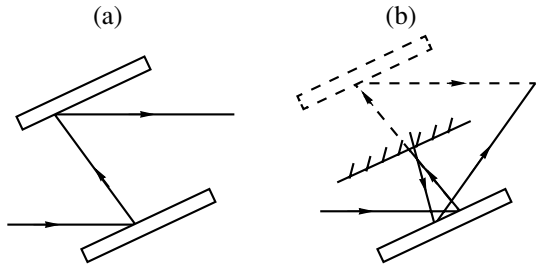

Fig. 3. (a) A two-grating system widely used in lasers to produce short light pulses [3]. (b) The path of the diffracted beams in the constant phase autocorrelation interferometer and in an equivalent two-grating system.

For a classical harmonic oscillator with a low damping γ, we assume, by disregarding phase modulation, that

$$\exp(i\psi)\equiv 1,\qquad A(t) = A_0\theta(t-t_{\text{on}})\exp[-\gamma(t-t_{\text{on}})], \tag{11}$$

where t_{on} is the time of oscillator excitation. The wave packet at the output takes the form

$$B_{\text{out}}(t'+\tau_0) = \left(\frac{\mu}{2\pi}\right)^{1/2}\exp\left[i\left(\phi_0-\frac{\pi}{4}-\omega_0(t'+\tau_0)\right)\right]A_0 \times\int_{-\infty}^{\infty}\theta(t-t_{\text{on}})\exp\left[-\gamma(t-t_{\text{on}})+i\frac{\mu}{2}(t-t')^2\right]dt. \tag{12}$$

For $\gamma = 0$ (undamped oscillator), calculating the integral in (12) is formally similar to the problem of plane-wave diffraction by a half-plane:

$$B_{\text{out}}(t'+\tau_0) = \left(\frac{\mu}{2\pi}\right)^{1/2}\exp\left[i\left(\phi_0-\frac{\pi}{4}-\omega_0(t'+\tau_0)\right)\right]A_0 \times\int_{t_{\text{on}}}^{\infty}\exp\left[i\frac{\mu}{2}(t-t')^2\right]dt = \frac{A_0}{\sqrt{2}}\exp\left[i\left(\phi_0-\frac{\pi}{4}-\omega_0(t'+\tau_0)\right)\right] \times\left[\frac{\exp(-i\pi/4)}{\sqrt{2}}+C\left(\sqrt{\frac{\mu}{\pi}}(t'-t_{\text{on}})\right)+iS\left(\sqrt{\frac{\mu}{\pi}}(t'-t_{\text{on}})\right)\right]. \tag{13}$$

Here, $C(x)$ and $S(x)$ are the Fresnel integrals. The analogy with Fresnel diffraction allows the distortion of the wave packet when it passes through the interferometer to be visualized. In this case, time variations in the light intensity (pulse envelope) are shown in Fig. 4. The light intensity reaches a maximum not at $t' = t_{\text{on}}$ but with a

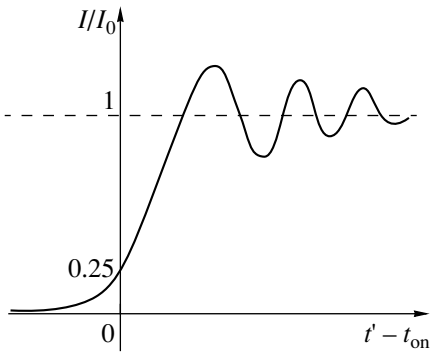

Fig. 4. Time variations in light intensity (pulse envelope) for an undamped harmonic oscillator switched on at time t_{on}.

delay introduced by the diffraction grating. Characteristic damped oscillations and a monotonic signal decay into the region of a time "shadow," $t' \le t_{on}$, are observed at the leading edge.

For a nonzero γ, the integral in (12) cannot be taken explicitly:

$$B_{out}(t' + \tau_0) = \left(\frac{\mu}{2\pi}\right)^{1/2} \times \exp\left[i\left(\phi_0 - \frac{\pi}{4} - \omega_0(t' + \tau_0)\right)\right] A_0 \times \exp\left[-\gamma(t' - t_{on}) + i\frac{\gamma^2}{2\mu}\right] J(t' - t_{on}, \gamma, \mu). \quad (14)$$

Here, we denoted

$$J(t' - t_{on}, \gamma, \mu) = \int_{-(t' - t_{on})}^{\infty} \exp\left[i\frac{\mu}{2}\left(z + \frac{\gamma}{\mu}i\right)^2\right] dz. \quad (15)$$

If the photodetector at the interferometer output responds to a time-integrated signal, then its response P will be given by the integral of the modulus of the total amplitude squared:

$$P = \int_{-\infty}^{\infty} |B_{in}(t'' + \tau_0) + B_{out}(t'')|^2 dt'' = \int_{-\infty}^{\infty} |B_{in}(t'' + \tau_0)|^2 dt'' + \int_{-\infty}^{\infty} |B_{out}(t'')|^2 dt'' + \int_{-\infty}^{\infty} 2\mathrm{Re}[B_{in}^*(t'' + \tau_0) B_{out}(t'')] dt''. \quad (16)$$

The interference term in (16) is

$$\int_{-\infty}^{\infty} 2\mathrm{Re}[B_{in}^*(t'' + \tau_0) B_{out}(t'')] dt'' = |A_0|^2 \left(\frac{\mu}{2\pi}\right)^{1/2} \times 2\mathrm{Re}\left\{\exp\left[i\left(\phi_0 - \frac{\pi}{4} + \frac{\gamma^2}{2\mu} + \omega_0\tau_0\right)\right] \times \int_{t_{on} - \tau_0}^{\infty} \exp[-\gamma(2(t'' - t_{on}) + \tau_0)] \times J(t'' - \tau_0 - t_{on}, \gamma, \mu) dt''\right\}. \quad (17)$$

To obtain signal W from an ensemble of oscillators, we must perform averaging over the carrier frequency ω_0 with the distribution function $f(\omega_0)$:

$$\int_0^{\infty} f(\omega_0) d\omega_0 = 1, \qquad W = \int_0^{\infty} P(\omega_0) f(\omega_0) d\omega_0. \quad (18)$$

Significantly, when the phase-constancy condition is satisfied, the carrier-wave phase $\omega_0\tau_0$ does not depend on ω_0; as a result, the phase factor may be taken outside the integral in (18). This implies that the interference pattern will not blur because of inhomogeneous broadening. The change in visibility (spectral luminous efficacy) of the interference pattern with separation between the mirror and the grating is given by integrals (17) and (18). By measuring the pattern visibility, we can determine parameters on which these integrals depend and, primarily, the homogeneous linewidth γ.

Let us now consider the effect of a finite grating resolution. Clearly, for a wave packet with a spectral width $\delta\omega$ ($\delta\lambda$ in wavelengths) smaller than the minimum spectral interval $\Delta\omega$ ($\Delta\lambda$ in wavelengths) allowable by the grating, the grating will act just as a mirror (with an additional time delay), and our interferometer will be similar to the Michelson interferometer. For an inhomogeneous broadening exceeding $\Delta\omega$, different wave packets will be directed by the grating along different paths and will finally produce the same interference pattern on the screen. In this case, the disappearance of the interference pattern will be determined by the minimum wavelength interval $\Delta\lambda$ allowable by the grating. For wave packets whose carrier wavelengths differ by no more than $\Delta\lambda$, only the first term should be retained in Eq. (4) when calculating the phase difference, because no change in the optical path caused by the grating action occurs for them. If we take $\Delta\Phi = -\pi$

as the disappearance condition for the interference pattern, then

$$\Delta\Phi = -2\pi\frac{S_{10}}{\lambda^2}\Delta\lambda = -2\pi\frac{N}{R} = -\pi \text{ or } R = 2N, \quad (19)$$

where $R = \lambda/\Delta\lambda$ is the grating resolution. In other words, the interference pattern disappears as the separation between the mirror and the grating increases when the resolution of the Michelson interferometer $2N$ reaches the grating resolution R. Let us write the order of interference N using (1):

$$R = 2\frac{S_{10}}{\lambda} = \frac{4z}{\lambda\cos\phi}(1 + \sin\phi\sin\theta). \quad (20)$$

Relation (20) clearly shows that as the grating resolution increases, the separation z at which the interference pattern ceases to be observable will increase proportionally. If we successively increase the grating resolution (for example, by replacing the grating with a new one with a large number of lines), then the separation z will increase until $\Delta\lambda = \delta\lambda$. This is the condition of equality between the homogeneous emission linewidth and the minimum wavelength interval allowable by the grating. As the resolution increases further, different Fourier components of each wave packet will propagate along different paths, and the wave packet will be distorted, as was described above; i.e., the grating will behave as a perfect one.

Let us estimate the spatial size of the wave packet for a weakly damped classical oscillator. Since the phase modulation for such an oscillator is small, we have

$$T\delta\omega \approx 1, \quad (21)$$

where T is the wave-packet duration. We define the spatial size L of the wave packet as

$$L = cT = \frac{c}{\delta\omega}. \quad (22)$$

Compare this size with the optical path difference S_{10}^* in our interferometer at which the frequency interval allowable by the grating becomes equal to the spectral width of the wave packet, $\Delta\omega = \delta\omega$:

$$R = \frac{\lambda}{\Delta\lambda} = \frac{\omega}{\Delta\omega} = \frac{\omega}{\delta\omega} = \frac{2S_{10}^*}{\lambda}, \text{ i.e., } L = \frac{S_{10}^*}{\pi}. \quad (23)$$

Thus, L and S_{10}^* are of the same order of magnitude. It may be roughly assumed that the interference pattern disappears when the path difference S_{10}^* is exceeded, because the wave packets from different interferometer arms do not overlap. To determine the specific law of reduction in visibility of the interference pattern with separation z between the grating and the mirror requires a detailed calculation and knowledge of the wave-packet shape. Once the grating reaches a spectral resolution that is enough to separate a single wave packet into its components, its autocorrelation analysis will essentially take place. So, carrying out appropriate experiments, it is hoped that the inverse problem—restoring the original shape of a single-photon wave packet—can be solved.

It is of interest to determine which wave packets can be subjected to an autocorrelation analysis with a grating of a given resolution. It follows from relation (23) that $L \le R\lambda/2\pi$. Using large gratings, we can reach $R = 200000$, which corresponds to $L < 1.6$ cm for a wavelength of 0.5 μm. The radiative lifetime corresponding to this distance, $T \approx 50$ ps, is an upper limit on the radiative lifetimes measurable with a given grating. Note that the lifetimes of interest in semiconductors are generally much shorter.

Thus, achieving a maximum grating resolution seems of great importance in our interferometer. The interferometer considered above (see Fig. 1), which is convenient for illustrating the basic principles, proves to be inefficient from this point of view. Indeed, traversing a finite, moderately large distance, the light diffracted by the grating and reflected by the mirror returns to the grating. As a result, not the entire grating but only the lines in the first Fresnel zone with an order-of-magnitude size $(2\lambda z)^{1/2}$ are effectively used.

To experimentally test the above considerations, the interferometer assembled according to the scheme in Fig. 1 was placed under an airtight cap, from which air was pumped out with a high-pressure vacuum pump. The air pressure under the cap was checked by a mercury pressure gauge. The beam from a helium-neon laser operated in TEM00 mode that was precollimated with a telescope until a divergence of 10^{-4} rad was directed toward a grating with 2400 lines/mm. The light emerging from the interferometer was collected by a long-focal-length objective ($f = 210$ mm) and recorded by a photodiode in its focal plane. Air was allowed to slowly bleed under the cap through a small hole. In this case, the photodiode recorded the periodic variations in light intensity attributable to the dependence of the refractive index for air on the pressure under the cap. The change in air refractive index as the pressure increased from 20 to 660 mm Hg caused the wavelength to change by about 1.6 Å. The number of passed interference maxima for a perfect grating with an infinite resolution is given by

$$\Delta N = \Delta\left(\frac{S_{10}}{\lambda_0}n_a\right) = \left(\frac{S_{10}}{\lambda_0} + \frac{n_a}{\lambda_0}\frac{dS_{10}}{d\lambda}\frac{d\lambda}{dn_a}\right)\Delta n_a = \left(\frac{S_{10}}{\lambda_0} - \frac{2z\lambda_0}{h^2 n_a^2\cos^3\phi}\right)\Delta n_a, \quad (24)$$

where λ_0 is the wavelength of the helium-neon laser in a vacuum, and n_a is the refractive index for air ($n_a \approx 1.0003$ under normal pressure). We see that it must be smaller than the number of maxima that passed in a similar situation in the Michelson interferometer, for

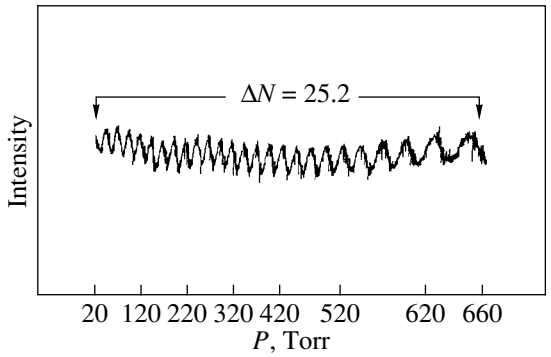

Fig. 5. Light intensity at the output of the interferometer assembled according to the scheme in Fig. 1 and placed under a pumped cap versus air pressure under the cap; z = 19 mm and $\theta = 75°$.

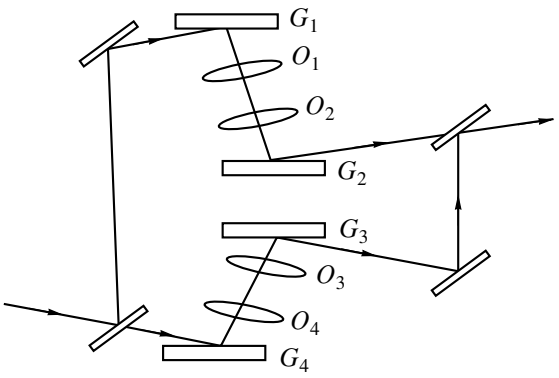

Fig. 6. An improved constant phase autocorrelation interferometer, which allow one to make full use of the grating resolution and to vary the arm difference over a wide range.

which only the first term is available in Eq. (24). The experimental curve shown in Fig. 5 actually demonstrates a reduction in the number of passed maxima attributable to the grating action. In the pressure range shown in Fig. 5, 27.9 interference orders must have been observed, while the experimental curve contains a mere 25.2 orders. Check measurements with a Michelson interferometer yielded excellent agreement between theory and experiment. Experiments with a lower-resolution (1200 lines per mm) grating show that the difference from the Michelson interferometer (with the same optical path difference) decreased to 0.6 of the interference order. Note that a nonoptimal use of the grating resolution causes the observed effect to be reduced considerably (by more than an order of magnitude) compared to an infinite resolution that corresponds to Eq. (24).

An improved interferometer can be proposed for practical measurements. This interferometer makes full use of the grating resolution and makes it possible to change the path difference between the interfering beams over a wide range. Its scheme is shown in Fig. 6. As in a Michelson interferometer, light is split by a semitransparent plate into two arms, each containing a system of two parallel diffraction gratings. An afocal optical system with a unit angular magnification between the gratings ensures the grating operation in parallel beams; as a result, the maximum possible grating resolution can be achieved (it is important that the minimum spectral interval allowable by the gratings be smaller than the homogeneous linewidth). The independent change in interferometer arm length allows the visibility of the interference pattern to be measured over a wide range of beam path differences.

ACKNOWLEDGMENTS

I am grateful to G.B. Lesovik, the numerous discussions with whom helped clearly formulate the problem. I also wish to thank V.D. Kulakovskiĭ for support, D.L. Alov for the opportunity to carry out experiments on a holographic table, and E.V. Lebedeva for help in computations.

REFERENCES

1. Y. R. Shen, *The Principles of Nonlinear Optics* (Wiley, New York, 1984; Nauka, Moscow, 1989).
2. L. A. Vasil'ev and I. V. Ershov, *Grating Interferometer* (Mashinostroenie, Moscow, 1976).
3. E. B. Treacy, IEEE J. Quantum Electron. **QE-5**, 454 (1969).

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 93, No. 4, 2001, pp. 723–736.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 830–845.

ATOMS, SPECTRA, RADIATION

Effects of the Velocity Dependence of the Collision Frequency on the Dicke Line Narrowing

A. I. Parkhomenko and A. M. Shalagin*

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

**e-mail: shalagin@iae.nsk.su*

Received May 22, 2001

Abstract—The effect of the velocity (v) dependence of the transport collision frequency ν_{trv} on the Dicke line narrowing is analyzed in terms of the strong-collision model generalized to velocity-dependent collision frequencies (the so-called kangaroo model). This effect has been found to depend on the mass ratio of the resonance (M) and buffer (M_b) particles, $\beta = M_b/M$: it is at a minimum for $\beta \ll 1$ and reaches a maximum for $\beta \gtrsim 3$. A power-law particle interaction potential, $U(r) \propto r^{-n}$, is used as an example to show that, compared to $\nu_{trv}(v) = \text{const}$ ($n = 4$), the line narrows if $\nu_{trv}(v)$ decreases with increasing v ($n < 4$) and broadens if $\nu_{trv}(v)$ increases with v ($n > 4$). At $\beta \gtrsim 3$, the line width can increase [compared to $\nu_{trv}(v) = \text{const}$] by 5 and 12% for the potentials with $n = 6$ and $n \gtrsim 10$, respectively; for the potentials with $n = 1$ (Coulomb potential) and $n = 3$, it can decrease by more than half and 6%, respectively. The line profile $I(\Omega)$ has been found to be weakly sensitive to $\nu_{trv}(v)$ at some detuning Ω_c of the radiation frequency Ω. Dicke line narrowing is used as an example to analyze the collisional transport of nonequilibrium in the resonance-particle velocity distribution in a laser field. The transport effect is numerically shown to be weak. This allows simpler approximate one-dimensional quantum kinetic equations to be used instead of the three-dimensional ones to solve spectroscopic problems in which it is important to take into account the velocity dependence of the collision frequency when the phase memory is preserved during collisions. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is well known that elastic collisions of gas particles with the phase of the radiation-induced dipole moment conserved result in line narrowing (Dicke effect) [1–3]. Physically, the essence of the narrowing is that collisions with phase memory do not cause any collisional line broadening but restrict the spatial particle motion, thereby removing inhomogeneous Doppler broadening.

The models of strong [2, 3] and weak [2–4] collisions are commonly used to describe the line shape with allowance for the Dicke narrowing due to elastic collisions (recall that the collision frequencies in these models are velocity-independent). Precise measurements of the line profiles in the infrared molecular absorption spectrum show that in several cases, these models faithfully describe the observed (narrowed) line profiles (see, e.g., [5] and references therein). However, in some cases (for example, for CO molecules in He, Ne, Ar, Xe, and N_2 buffer gases [6], C_2H_2 molecules in a Xe buffer gas [7], and HF molecules in an Ar buffer gas [8]), none of these models gives satisfactory agreement with experiment. Reasonable agreement is achieved only when the Dicke narrowing due to elastic collisions (in the strong- and weak-collision models) and the dependences of the collisional line width and shift on the molecular velocity are simultaneously taken into account (see [6–8] and references therein). Nevertheless, difficulties in describing the experimentally observed absorption-line asymmetry, for example, for HF molecules in an Ar buffer gas [8], also arise in this approach. Accordingly, in order to adequately interpret the observed anomalies, Pine [8] suggested that, apart from the velocity dependence of the line width and shift, the velocity dependence of the collision frequency was also taken into account when describing the Dicke line narrowing. No formula was derived in [8] for the line profile. In essence, it was suggested describing the line profile by the well-known formula of the strong-collision model [2, 3] by substituting velocity-dependent frequencies for the velocity-independent collision frequencies in it. Of course, this substitution is physically incorrect, because it makes the Maxwell velocity distribution a nonequilibrium one, as was pointed out by Pine [8]. Good agreement with experimental results was considered to be a criterion for the applicability of the formula for the line profile proposed in [8]. Pine [8] drew attention to the necessity of solving the corresponding theoretical problem in an effort to properly allow for the velocity dependence of the collision frequency when calculating the Dicke narrowing.

In [9–11], it was theoretically shown that the velocity dependence of the collision frequency significantly

affects the Dicke line narrowing. In [9, 10], the absorption line of ions in equilibrium plasma was calculated by using the Landau collision integral. In [11], the Dicke narrowing was calculated by using the kernel of the collision integral in the hard-sphere model.

Thus, the importance of taking into account the velocity dependence of the collision frequency when calculating the Dicke line narrowing is beyond question. Accordingly, it is of interest to consider the Dicke effect in terms of a universal collision model that would allow the influence of the velocity dependence of the collision frequency on the Dicke narrowing to be described for any interaction potential of the colliding particles and at any ratio of their masses. It seems that the so-called kangaroo model [12, 13], which is a generalization of the strong-collision model to velocity-dependent collision frequencies, may be taken as such a universal collision model.

Here, our goal is to analyze the effect of the velocity dependence of the collision frequency on the Dicke line narrowing in terms of the kangaroo model.

2. GENERAL RELATIONS

The profile of a single spectral line, $I(\Omega)$, normalized in area to unity,

$$\int_{-\infty}^{\infty} I(\Omega)d\Omega = 1,$$

is given by

$$I(\Omega) = \frac{1}{\pi}\mathrm{Re}\langle\rho\rangle, \quad \langle\rho\rangle = \int\rho(\mathbf{v})d\mathbf{v}, \qquad \Omega = \omega - \omega_{mn}, \tag{1}$$

where ω and ω_{mn} are the field frequency and the Bohr m–n transition frequency, respectively; $\mathbf{v}$ is the resonance-particle velocity; and $\rho(\mathbf{v})$ is the nondiagonal density matrix element. For a low radiation intensity, $\rho(\mathbf{v})$ can be calculated from the kinetic equation (see, e.g., [3])

$$[\gamma - i(\Omega - \mathbf{k}\cdot\mathbf{v})]\rho(\mathbf{v}) = S(\mathbf{v}) + W(\mathbf{v}). \tag{2}$$

Here, $\gamma = (\Gamma_m + \Gamma_n)/2$, Γ_m and Γ_n are the spontaneous decay rates of the combining (radiation-affected) levels m and n; $\mathbf{k}$ is the radiation wave vector; $S(\mathbf{v})$ is the nondiagonal collision integral; and $W(\mathbf{v})$ is the Maxwell velocity distribution.

The general expression for the collision integral is

$$S(\mathbf{v}) = -\nu(v)\rho(\mathbf{v}) + \int A(\mathbf{v}|\mathbf{v}')\rho(\mathbf{v}')d\mathbf{v}', \tag{3}$$

where $\nu(v)$ is the "departure" frequency for the nondiagonal density matrix element, and $A(\mathbf{v}|\mathbf{v}')$ is the kernel of the collision integral. The actual kernel must satisfy the relation [3, 14]

$$A(\mathbf{v}'|\mathbf{v})W(\mathbf{v}) = A(\mathbf{v}|\mathbf{v}')W(\mathbf{v}'). \tag{4}$$

It thus follows that

$$\int A(\mathbf{v}|\mathbf{v}')W(\mathbf{v}')d\mathbf{v}' = \tilde{\nu}(v)W(\mathbf{v}), \qquad \tilde{\nu}(v) \equiv \int A(\mathbf{v}'|\mathbf{v})d\mathbf{v}', \tag{5}$$

where $\tilde{\nu}(v)$ is the "arrival" frequency [3, 14] for the nondiagonal density matrix element. In general, the velocity-dependent departure $\nu(v)$ and arrival $\tilde{\nu}(v)$ frequencies are complex-valued. The relation

$$\gamma + \nu(v) - \tilde{\nu}(v) = \Gamma_1(v) + i\Delta_1(v) \tag{6}$$

gives the absorption-line half-width $\Gamma_1(v)$ and shift $\Delta_1(v)$ for particles with a fixed velocity v without Doppler broadening [3, 14].

The kangaroo model [12, 13] is based on the assumption that the kernel of the collision integral is factorized:

$$A(\mathbf{v}|\mathbf{v}') = f_1(\mathbf{v})f_2(\mathbf{v}'). \tag{7}$$

The following expression can be derived from (5) and (7) for the kernel of the collision integral in the kangaroo model:

$$A(\mathbf{v}|\mathbf{v}') = \frac{\tilde{\nu}(v)W(\mathbf{v})\tilde{\nu}(v')}{\langle\tilde{\nu}\rangle}, \qquad \langle\tilde{\nu}\rangle = \int\tilde{\nu}(v)W(\mathbf{v})d\mathbf{v}. \tag{8}$$

Substituting $A(\mathbf{v}|\mathbf{v}')$ from (8) in (3) yields an expression for the nondiagonal collision integral in the kangaroo model:

$$S(\mathbf{v}) = -\nu(v)\rho(\mathbf{v}) + \frac{\tilde{\nu}(v)W(\mathbf{v})}{\langle\tilde{\nu}\rangle}\int\tilde{\nu}(v)\rho(\mathbf{v})d\mathbf{v}. \tag{9}$$

The kangaroo model (9) is a generalization of the strong-collision model to velocity-dependent collision frequencies. For velocity-independent collision frequencies, the kangaroo model (9) transforms to the strong-collision model.

The following expression can be derived from (1), (2), and (9) for the line profile:

$$I(\Omega) = \frac{1}{\pi k v_T}\mathrm{Re}\left\{Y_0 + \frac{\frac{\langle\tilde{\nu}\rangle}{kv_T}Y_1^2}{1 - \frac{\langle\tilde{\nu}\rangle}{kv_T}Y_2}\right\}, \tag{10}$$

where

$$Y_q = \int\frac{kv_T W(\mathbf{v})}{\gamma + \nu(v) - i(\Omega - \mathbf{k}\cdot\mathbf{v})}\left(\frac{\tilde{\nu}(v)}{\langle\tilde{\nu}\rangle}\right)^q d\mathbf{v}, \qquad v_T = \sqrt{\frac{2k_BT}{M}}, \quad q = 0, 1, 2, \tag{11}$$

M and v_T are the mass and most probable velocity of the absorbing particles, respectively; T is the temperature; and k_B is the Boltzmann constant.

Formula (10) differs in structure from the standard formula for the line profile in the strong-collision model [2, 3]. The well-known expression for the line profile in the strong-collision model [2, 3, 14] follows from (10) for $\tilde{\nu}(v)$, $\nu(v) = \text{const}$.

Having performed the integration over the directions of velocity $\mathbf{v}$ in (8) and (11), we derive for $\langle\tilde{\nu}\rangle$ and Y_q in (10)

$$\langle\tilde{\nu}\rangle = \frac{4}{\sqrt{\pi}}\int_0^\infty t^2\exp(-t^2)\tilde{\nu}(tv_T)dt,$$

$$\begin{aligned}\mathrm{Re}(Y_q) &= \frac{2}{\sqrt{\pi}}\int_0^\infty t\exp(-t^2)\psi(t)\left(\frac{\tilde{\nu}(tv_T)}{\langle\tilde{\nu}\rangle}\right)^q dt,\\ \mathrm{Im}(Y_q) &= -\frac{1}{\sqrt{\pi}}\int_0^\infty t\exp(-t^2)L(t)\left(\frac{\tilde{\nu}(tv_T)}{\langle\tilde{\nu}\rangle}\right)^q dt,\end{aligned} \qquad (12)$$

$$q = 0, 1, 2.$$

Here, we introduced the functions of dimensionless velocity $t = v/v_T$

$$\psi(t) = \arctan\frac{t+x(t)}{y(t)} + \arctan\frac{t-x(t)}{y(t)},$$

$$L(t) = \ln\left[\frac{y^2(t)+[t-x(t)]^2}{y^2(t)+[t+x(t)]^2}\right], \qquad (13)$$

$$y(t) = \frac{\gamma + \mathrm{Re}[\nu(tv_T)]}{kv_T}, \quad x(t) = \frac{\Omega - \mathrm{Im}[\nu(tv_T)]}{kv_T}.$$

Thus, calculating the line profile in the kangaroo model reduces to calculating single integrals.

3. THE CASE OF $\tilde{\nu}(v) = \nu(v)$

If the effects of phase memory in collisions may be disregarded (the elastic scattering amplitudes of the particles in the combining states m and n differ significantly), then we should set $\tilde{\nu}(v) = 0$ in the kangaroo model (9). According to (6), the quantity $\gamma + \nu(v)$ is determined by the line half-width $\Gamma_1(v)$ and shift $\Delta_1(v)$ [as a rule, the radiation half-width γ is known; in general, $\gamma \ll \nu(v)$ for molecules]. In this case, the line profile is described by the function $\mathrm{Re}(Y_0)$.

If, alternatively, the phase memory is completely preserved in collisions (the elastic scattering amplitudes of the particles in the combining states m and n are equal; the most favorable conditions for the Dicke narrowing to show up), then, as is well known [3, 14], the nondiagonal kernel $A(\mathbf{v}|\mathbf{v}')$ is real and matches the diagonal kernel $A_i(\mathbf{v}|\mathbf{v}')$ of the collision integral for the resonance particles in state $i = m, n$:

$$A(\mathbf{v}|\mathbf{v}') = A_n(\mathbf{v}|\mathbf{v}') = A_m(\mathbf{v}|\mathbf{v}').$$

In this case, the departure $\nu(v)$ and arrival $\tilde{\nu}(v)$ frequencies are also real and equal [3, 14]:

$$\tilde{\nu}(v) = \nu(v). \qquad (14)$$

The subsequent analysis of the line profile (10) is performed for (14), when the phase memory is completely preserved during collisions; the collision frequency (14) is the only model parameter.

Formulas (10)–(13) give the solution of our problem in quadratures for any dependence $\nu(v)$. However, there is little point in discussing the abstractly postulated dependences $\nu(v)$. It is of interest to study the effect of the mass ratio of the colliding particles and a specific interaction potential on $\nu(v)$ and, hence, on the line profile $I(\Omega)$. To this end, it makes sense to trace the relationship of $\nu(v)$ to the characteristics of an elementary scattering event. Given that $\nu(v)$ is the departure frequency and that it is generally expressed in terms of the total scattering cross section [3, 14], one would think that this relationship can be clearly established. In reality, this issue must be addressed with caution: when the collision integral is modeled, the physical treatment of the model parameters can change somewhat. In particular, in the kangaroo model used here, $\nu(v)$ in conditions (14) also acts as the so-called transport collision frequency $\nu_{trv}(v)$, which enters into the coefficient of proportionality between the drag force and the specified particle velocity:

$$\begin{aligned}\mathbf{F} &= -M\nu_{\mathrm{tr}v}(v)\mathbf{v},\\ \nu_{\mathrm{tr}v}(v) &\equiv \int\left(1-\frac{\mathbf{v}\cdot\mathbf{v}_1}{v^2}\right)A(\mathbf{v}_1|\mathbf{v})d\mathbf{v}_1.\end{aligned} \qquad (15)$$

The transport frequency $\nu_{\mathrm{tr}v}(v)$ is expressed in terms of the transport scattering cross section [14], and its velocity dependence differs from that for the departure frequency.

Thus, as a parameter of the kangaroo model, $\nu(v)$ also simultaneously acts as the departure and transport frequencies. It should be realized with what the model parameter $\nu(v)$ must be associated when the Dicke problem is considered. This question can be answered by analyzing the asymptotic value of $I(\Omega)$ at large buffer-gas pressures. When the conditions $\langle\tilde{\nu}\rangle \gg kv_T$,

γ, $|\Omega|$ are satisfied, the line profile $I(\Omega)$ is described by the formula

$$I(\Omega) = \frac{1}{\pi}\frac{\gamma+\gamma_{tr}}{(\gamma+\gamma_{tr})^2+\Omega^2},$$
$$\gamma_{tr} = k^2\int\frac{(\mathbf{n}\cdot\mathbf{v})^2 W(\mathbf{v})}{\nu(v)}d\mathbf{v} \qquad (16)$$
$$= \frac{4k^2}{3\sqrt{\pi}v_T^3}\int_0^\infty\frac{v^4}{\nu(v)}\exp\left(-\frac{v^2}{v_T^2}\right)dv.$$

Here, $\mathbf{n}$ is a unit vector in an arbitrarily chosen direction.

Under the same conditions, the line profile can formally be calculated in the general case without modeling the collision integral (see the Appendix). It is described by the same universal formula (16), where $1/\tau(v)$ acts as $\nu(v)$ [see formulas (A.7) and (A.8)]. The quantity $\tau(v)$ is a transport parameter and can be interpreted as the time it takes to lose memory of the initial particle velocity direction due to collisions. In general, $\tau(v)$ is unrelated to the parameters of an elementary scattering event, but it is clear that $1/\tau(v)$ is physically much closer to the transport collision frequency than to the departure frequency. Thus, the kangaroo-model parameter $\nu(v)$ in the Dicke problem is physically much closer to the transport collision frequency than to the departure frequency. Below, we therefore ascribe the velocity dependence characteristic of the transport collision frequency to $\nu(v)$. An additional argument for this is that in the Lorentz gas model (the limiting case of heavy buffer particles), the diffusivity D, which γ_{tr} is proportional to ($\gamma_{tr} = k^2D$; see the Appendix), is expressed in terms of the transport collision frequency (15) in the same way as in (16) [15, 16].

An important parameter of the line profile $I(\Omega)$ is its half-width, Γ_w. Plots of Γ_w against buffer-gas density are usually analyzed when discussing the Dicke effect. For velocity-dependent collision frequencies, the behavior of these plots and their relative positions begin to significantly depend on precisely which physical quantity proportional to the density is laid off along the horizontal axis [the density proper, the value of $\langle\nu\rangle$, or the value of $\nu(v)$ averaged by a different method, the transport collision frequency averaged in one way or another, etc. may be laid off along this axis]. The corresponding plots for different mass ratios of the colliding particles and for different interaction potentials turn out to be spaced most closely if ν_{tr}, which is inversely proportional to γ_{tr}, is taken as the argument for Γ_w:

$$\nu_{tr} \equiv \frac{(kv_T)^2}{2\gamma_{tr}}$$
$$= \frac{v_T^2}{2\int(\mathbf{n}\cdot\mathbf{v})^2\tau(v)W(\mathbf{v})d\mathbf{v}} = \frac{v_T^2}{2D}, \qquad (17)$$

where D is the true diffusivity. In the kangaroo model, the diffusivity is expressed in terms of $\nu(v)$ according to (16). Note that the diffusivity introduced here is responsible for the corresponding transport in gas-kinetic problems (see the Appendix). As we see, the well-known gas-kinetic parameter also universally describes the line shape in the Dicke effect at asymptotically high buffer-gas pressures.

Thus, in order to analyze the pressure dependence of the line half-width $I(\Omega)$, it makes sense to consider the function $\Gamma_w(\nu_{tr})$. It has the same model-independent asymptotics (16) for any mass ratios and interaction potentials of the colliding particles. The quantity ν_{tr} itself (mean transport collision frequency) is directly related to a well-known physical parameter (diffusivity); it can be obtained either from tabulated data or from independent measurements of the diffusivity in transport phenomena.

Let us analyze the line profile $I(\Omega)$ by taking into account the above remarks. We will ascribe the velocity dependence characteristic of the transport frequency $\nu_{trv}(v)$ to $\nu(v)$. For the Maxwell velocity distribution of buffer particles, the transport collision frequency is related to the parameters of an elementary scattering event by [15]

$$\nu_{trv}(v) = \frac{q}{v^3}\int_0^\infty u^2\exp\left(-\frac{u^2+v^2}{\bar{v}_b^2}\right)F(uv)\sigma_{tr}(u)du, \qquad (18)$$

where

$$F(uv) = \frac{2uv}{\bar{v}_b^2}\cosh\left(\frac{2uv}{\bar{v}_b^2}\right) - \sinh\left(\frac{2uv}{\bar{v}_b^2}\right),$$
$$q = \frac{\mu}{M}\frac{N_b\bar{v}_b}{\sqrt{\pi}}, \quad \mu = \frac{MM_b}{M+M_b}, \quad \bar{v}_b = \sqrt{\frac{2k_BT}{M_b}}, \qquad (19)$$

N_b and M_b are the buffer-particle density and mass, u is the relative velocity of the resonance and buffer particles, and $\sigma_{tr}(u)$ is the transport cross section for the scattering of an absorbing particle by a buffer particle.

For a power-law particle interaction potential,

$$U(r) \propto r^{-n} \qquad (20)$$

in the classical limit, the transport scattering cross section $\sigma_{tr}(u)$ depends on the particle relative velocity u as [17][1]

$$\sigma_{tr}(u) \propto u^{-4/n}. \qquad (21)$$

[1] For the Coulomb particle interaction ($n = 1$), a logarithmic divergence arises when calculating $\sigma_{tr}(u)$. For charged particles in plasma, this divergence is known to be eliminated by cutting off the impact parameter at a distance of the order of the Debye screening length R_D.

For this $\sigma_{tr}(u)$, it follows from (18) that [14, 15]

$$\nu_{trv}(v) = \nu_{trv}(0)\,{}_1F_1\left(a; \frac{5}{2}; -\beta\frac{v^2}{v_T^2}\right), \qquad a \equiv \frac{2}{n} - \frac{1}{2}, \quad \beta = \frac{M_b}{M}, \tag{22}$$

where ${}_1F_1(a; b; x)$ is Kummer's degenerate hypergeometric function. Formula (22) can also be written as

$$\nu_{trv}(v) = \nu_{tr}^{(c)}(1+\beta)^a\,{}_1F_1\left(a; \frac{5}{2}; -\beta\frac{v^2}{v_T^2}\right), \tag{23}$$

where $\nu_{tr}^{(c)}$ is the mean transport collision frequency, which is related by

$$D^{(c)} = v_T^2/2\nu_{tr}^{(c)}$$

to the diffusivity $D^{(c)}$ calculated in the first approximation of Chapman–Enskog's method (or in the five-moment approximation of Grad's method). The quantity $\nu_{tr}^{(c)}$ is given by (see, e.g., [14])

$$\nu_{tr}^{(c)} = \frac{8}{3\sqrt{\pi}v_T^5}\int_0^\infty v^4 \exp\left(-\frac{v^2}{v_T^2}\right)\nu_{trv}(v)\,dv. \tag{24}$$

The parameter $\nu_{tr}^{(c)}$ (24) is equal to ν_{tr} (17) only for $\nu_{trv}(v) = \text{const}$, i.e., either when $\beta \longrightarrow 0$ (the limiting case of light buffer particles) or at $n = 4$ for a power-law interaction potential. For a power-law potential at $n \neq 4$, the following relation holds:

$$\nu_{tr}^{(c)} > \nu_{tr}.$$

The difference between ν_{tr} and $\nu_{tr}^{(c)}$ is at a maximum for heavy buffer particles ($\beta \gg 1$), and it is about 13% for the potentials with $n = 2$ (charge-dipole interaction for distances that are sufficiently large compared to the characteristic particle sizes) and $n = \infty$ (an equivalent of the hard-sphere model); for $2 < n < \infty$, this difference is even smaller.[2] For the Coulomb potential ($n = 1$), the difference between $\nu_{tr}^{(c)}$ and ν_{tr} is at a maximum, and the $\nu_{tr}^{(c)}/\nu_{tr}$ ratio reaches ≈ 3.4 at $\beta \gg 1$.

It follows from (23) that the transport collision frequency decreases with increasing v at $n < 4$ and increases with v at $n > 4$. At $n = 4$, the collision frequency is velocity-independent. For light buffer particles ($\beta \ll 1$), the frequency $\nu_{trv}(v)$ depends weakly on v. The v dependence of ν_{trv} increases with β and reaches its maximum for heavy buffer particles ($\beta \gg 1$).

When condition (14) is satisfied, $\mathrm{Im}[\nu(v)] = 0$, so the function $I(\Omega)$ reaches its maximum at $\Omega = 0$. Given (23), it follows from (10) for large Doppler broadening ($kv_T \gg \nu_{tr}, \gamma$) that at $\beta \ll 1$ (light buffer particles)

$$\sqrt{\pi}kv_T I(0) = 1 + \frac{(\pi-2)\left[1 - \dfrac{2(n-4)\beta}{5n}\right]\nu_{tr} - 2\gamma}{\sqrt{\pi}kv_T}, \tag{25}$$

while at $\beta \gg 1$ (heavy buffer particles) and $n \geq 4$

$$\sqrt{\pi}kv_T I(0) = 1 - \frac{2\gamma}{\sqrt{\pi}kv_T} + \frac{\nu_{tr}}{\sqrt{\pi}kv_T} \times \frac{4n^2\Gamma\left(3-\frac{2}{n}\right)\Gamma\left(2+\frac{2}{n}\right)}{3\pi(n-1)(n-2)}\left\{\frac{\pi^2(n-2)}{4n}\left[\frac{\Gamma\left(\frac{3}{2}-\frac{2}{n}\right)}{\Gamma\left(2-\frac{2}{n}\right)}\right]^2 - 1\right\}, \tag{26}$$

where $\Gamma(x)$ is the Gamma function. For the interaction potential with $n = 4$, formulas (25) and (26) transform to a well-known formula for the central line intensity in the strong-collision model [2, 3, 14], as must be the case because the transport frequency $\nu_{trv}(v) = \nu_{tr}$ does not depend on v at $n = 4$.

As follows from (25) and (26), the smaller n, the larger $I(0)$. Since the area under the line profile is constant and normalized to unity, an increase in $I(0)$ implies a decrease in the line width and vice versa. Thus, at the same ν_{tr}, the line width increases with n. This result is easiest to interpret in terms of the so-called one-dimensional approach [3, 13, 14] to solving spectroscopic problems in which it is important to take into account the velocity dependence of the collision frequency. This approach is considered in the next section.

In Figs. 1 and 2, the line half-widths Γ_w for the profile $I(\Omega)$ calculated using formula (10) are plotted against mean transport elastic collision frequency ν_{tr} at various n for a power-law interaction potential and at various mass ratios β of the resonance and buffer particles.

We clearly see from Fig. 1 that at the same ν_{tr}, the line width increases with n for a power-law potential. Allowance for the dependence $\nu_{trv}(v)$ increases the calculated line width compared to $\nu_{trv}(v) = \text{const}$ ($n = 4$, the strong-collision model) for $n > 4$ and decreases it for $n < 4$. As we see from Fig. 1a, allowance for $\nu_{trv}(v)$ at $\nu_{tr} < kv_T$ can increase the calculated line width [compared to $\nu_{trv}(v) = \text{const}$] by 5 and 12%, respectively, for the potentials with $n = 6$ and $n = \infty$ (the interaction

[2] While on the subject of a power-law potential with $n = \infty$, we have in mind that the transport cross section is $\sigma_{tr}(u) = \text{const}$ for $n = \infty$, as follows from (21). The velocity independence of the transport cross section is characteristic of the particle interaction in the hard-sphere model.

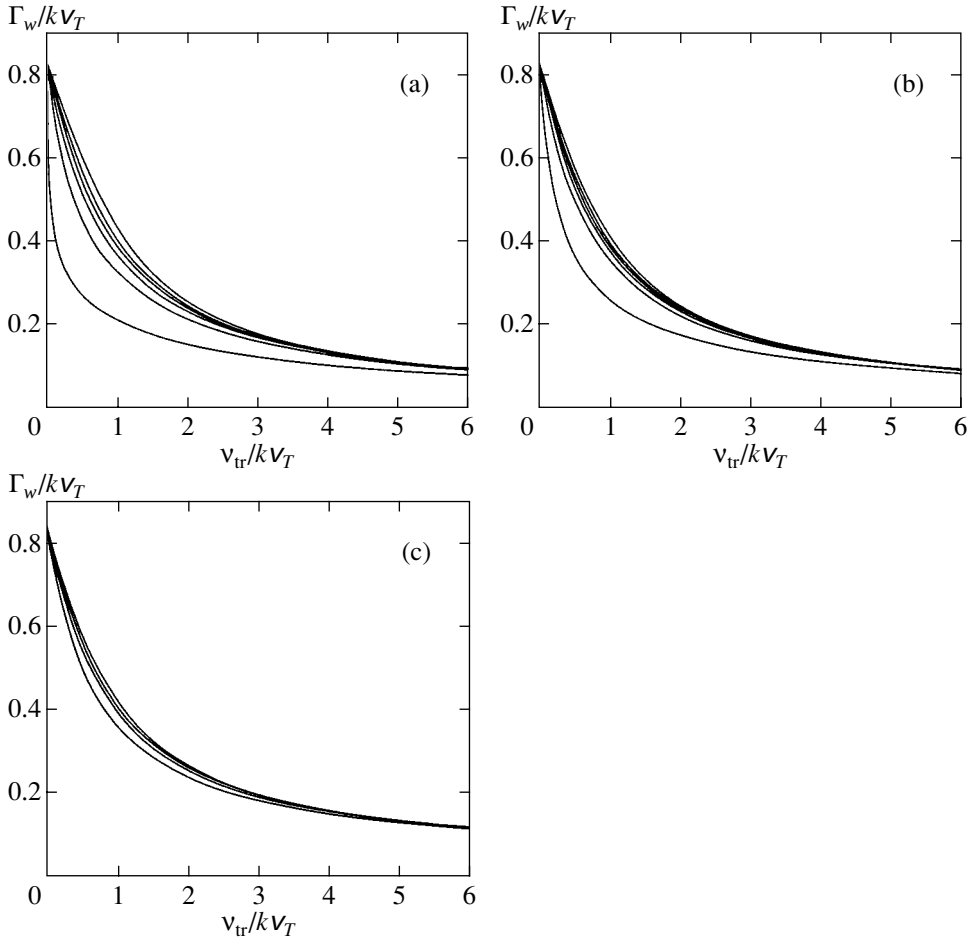

Fig. 1. Line half-width Γ_w versus mean transport collision frequency ν_{tr} at various n for a power-law interaction potential and at a fixed mass ratio β of the resonance and buffer particles. The values of n correspond to the arrangement of curves from top to bottom; $\gamma/kv_T = 10^{-2}$; (a) $\beta = 10$ and $n = \infty$, 6, 4, 3, 2, 1; (b) $\beta = 1$ and $n = \infty$, 6, 4, 3, 2, 1; and (c) $\beta = 0.3$ and $n = \infty$, 4, 2, 1.

potential for hard spheres). For the potential with $n = 3$, allowance for $\nu_{trv}(v)$ at $\nu_{tr} < kv_T$ can decrease the calculated line width by 6%. The dependence $\nu_{trv}(v)$ affects the calculated line width most strongly for the Coulomb interaction potential ($n = 1$). A comparison of the curves in Fig. 1a for $n = 1$ and $n = 4$ at $\nu_{tr}/kv_T \sim 0.5$ shows that allowance for $\nu_{trv}(v)$ can decrease the calculated line width by more than a factor of 2 compared to $\nu_{trv}(v) = \text{const}$.

The effect of the velocity dependence of the collision frequency on the line profile $I(\Omega)$ is at a minimum for light buffer particles, $\beta \ll 1$ (in this case, $\nu_{trv}(v)$ depends weakly on v), and at a maximum for heavy buffer particles ($\beta \gg 1$). A numerical analysis indicates that for the problem under consideration, the limit $\beta \gg 1$ is reached beginning with $\beta \approx 3$. In other words, $\beta = 3$ is virtually equivalent to the condition $\beta \gg 1$. Thus, the plots of line half-widths Γ_w against ν_{tr} at $\beta = 3$ are close to those in Fig. 1a for $\beta = 10$.

According to formula (16), the line half-width Γ_w in the limiting case of large collision frequencies ($\nu_{tr} \gg kv_T, \gamma$) does not depend on the interaction potential of the colliding particles and is $\Gamma_w = \gamma + \gamma_{tr}$. It follows from an analysis of the plots in Fig. 1a that the formula

$$\Gamma_w = \gamma + \gamma_{tr}$$

gives the line half-width with a satisfactory accuracy (no lower than 2% for $n \geq 3$) even at $\nu_{tr} \approx 5kv_T$ at any n for power-law interaction potentials except $n = 1$.

Figure 2 illustrates the effect of the mass ratio β of the resonance and buffer particles on the line half-width. The difference between the line half-widths for heavy ($\beta \gg 1$) and light ($\beta \ll 1$) buffer particles and, other things being equal, for the potentials with $n = 3$, 6, and ∞ reaches 6, 5, and 12%, respectively, while for the Coulomb potential ($n = 1$), this difference can be twice as large.

Figure 3 shows the line profiles $I(\Omega)$ at various n for a power-law interaction potential and mean transport collision frequency ν_{tr}. A characteristic feature of this

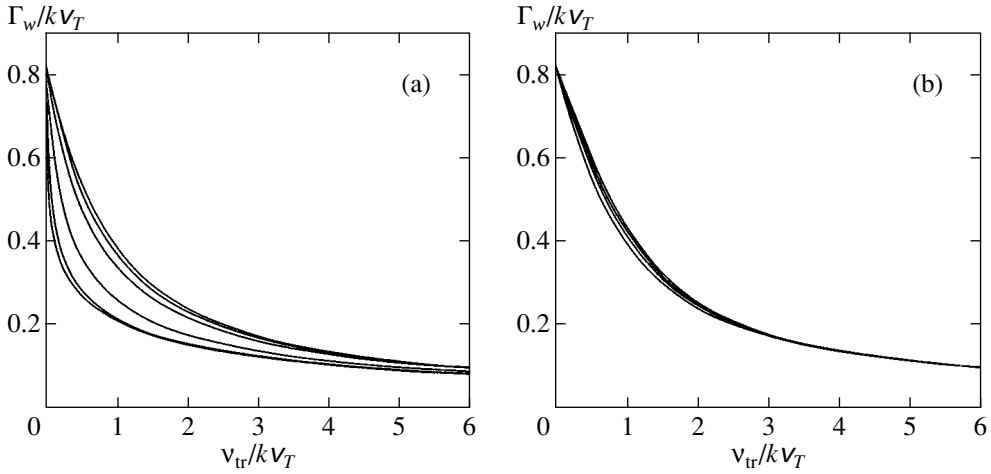

Fig. 2. Line half-width Γ_w versus mean transport collision frequency ν_{tr} at various mass ratios β of the resonance and buffer particles and at a fixed n for a power-law interaction potential. The values of β correspond to the arrangement of curves from top to bottom; $\gamma/kv_T = 10^{-2}$; (a) $n = 1$ and $\beta = 0.01, 0.1, 0.3, 1, 3, 100$; (b) $n = \infty$ and $\beta = 100, 3, 1, 0.1$.

figure is the intersection of all $I(\Omega)$ plots roughly at a single point for some values of the frequency detuning $\Omega = \Omega_c$ (for example, at $|\Omega|/kv_T \approx 0.33$ and 1.3 in Fig. 3a). When passing through the points $\Omega = \Omega_c$, the values of $I(\Omega)$ are "inverted": the largest value of $I(\Omega)$ (depending on n) becomes smallest and vice versa. The velocity dependence of the collision frequency weakly affects the intensity $I(\Omega_c)$ at these points.

Note the following. We analyzed the line width and shape as a function of the mean transport collision frequency ν_{tr}, which is related to the true, experimentally measurable diffusivity D by formula (17). In principle, we could also plot the half-width Γ_w against mean transport collision frequency $\nu_{tr}^{(c)}$ (24) or against $\langle\tilde{\nu}\rangle/kv_T$, which naturally arises in formula (10). A direct comparison of the behavior of the line half-width Γ_w as a function of these variables shows that the dependence of Γ_w on collision frequency (ultimately, on buffer-gas pressure) is most universal when ν_{tr} (17) is chosen as a variable. The plot of Γ_w against ν_{tr} rapidly (at $\nu_{tr} \gtrsim 5kv_T$) approaches the asymptotics $\Gamma_w = \gamma + \gamma_{tr}$ [see formula (16)] for any interaction potential. If the transport collision frequency $\nu_{trv}(v)$ decreases with increasing velocity, then the v dependence of ν_{trv} causes the line to narrow compared to the case where $\nu_{trv}(v) = \text{const}$. If, alternatively, $\nu_{trv}(v)$ increases with v ($n > 4$ for a power-law interaction potential), then the v dependence of ν_{trv} causes the line to broaden.

When plotting Γ_w against $\nu_{tr}^{(c)}$ or $\langle\tilde{\nu}\rangle$, neither a rapid approach to the asymptotics nor a universal law of line narrowing or broadening depending on whether ν_{trv} decreases of increases with v are observed. For example, Γ_w depends on $\nu_{tr}^{(c)}$ for power-law potentials with $n < 4$ as follows: the line narrows for $\nu_{tr}^{(c)} < kv_T$ and broadens beginning with $\nu_{tr}^{(c)} > kv_T$ compared to $\nu_{trv}(v) = \text{const}$.

We now draw attention to the following interesting fact. Since the collision model we use describes pair collisions of particles with each other, the possibility of applying our formulas to the Coulomb interaction of particles ($n = 1$) in plasma is called into question. The point is that the Coulomb interaction of charged particles with each other in plasma is a long-range interaction and is of a collective nature rather than a pair one. Nevertheless, a direct comparison of our calculations with the results from [9, 10], in which the Dicke absorption-line narrowing for ions in equilibrium plasma was calculated by using the Landau collision integral for $\beta = 1$, shows that the case $n = 1$ can also be satisfactorily described by our formulas.

Let us consider this case in more detail. At $n = 1$, the transport scattering cross section $\sigma_{tr}(u)$ in (18) can be calculated using Rutherford's formula for the effective scattering cross section of charged point particles. A logarithmic divergence arises in this case, which is associated with the large contribution of distant collisions that produce small-angle scattering. For plasma charged particles, this divergence is known (see, e.g., [18]) to be eliminated by cutting off the impact parameter ρ at a distance of the order of the Debye screening length R_D; i.e., the contribution from collisions with $\rho > R_D$ is disregarded. Substituting the transport scattering cross section $\sigma_{tr}(u)$ thus calculated in (18) yields ($t = v/v_T$)

$$\nu_{trv}(v) = \nu_{trv}(0)\frac{3}{2\beta t^2} \times\left[\frac{\sqrt{\pi}\,\mathrm{erf}(\sqrt{\beta}t)}{2\sqrt{\beta}t} - \exp(-\beta t^2)\right], \tag{27}$$

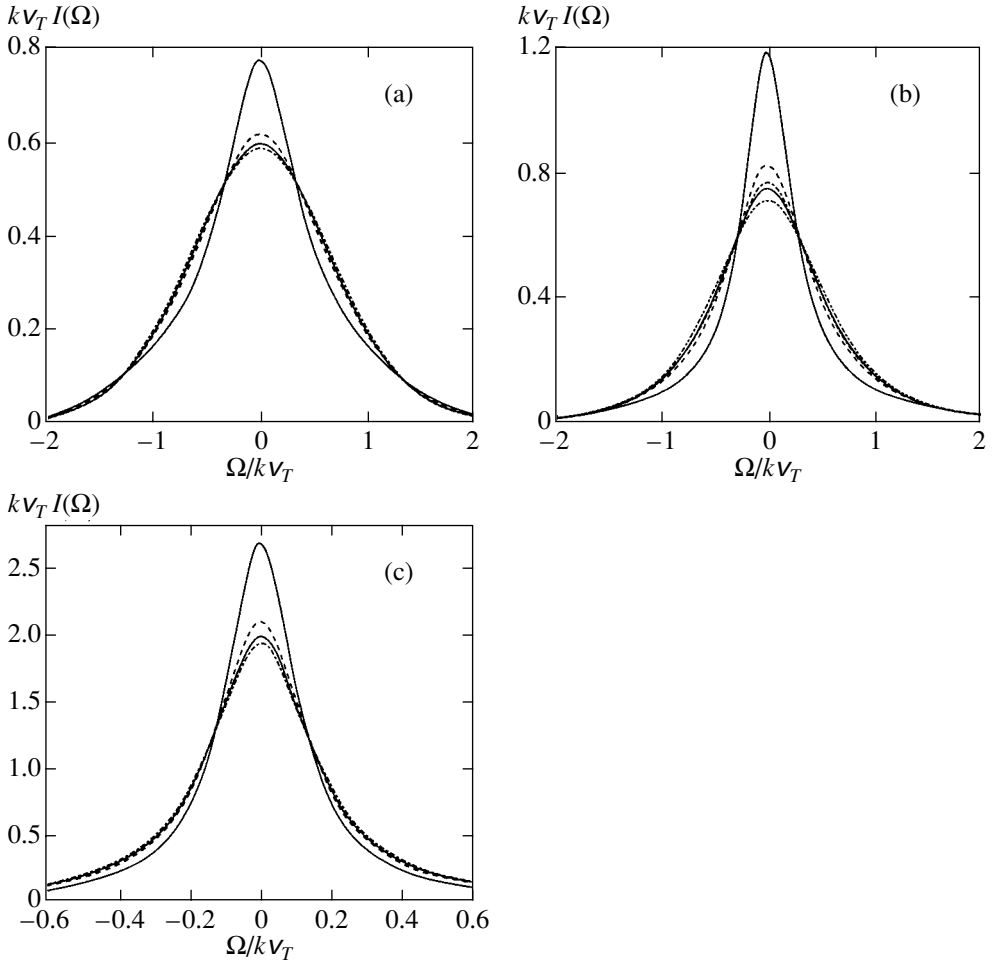

Fig. 3. Line profile $I(\Omega)$.The values of n for a power-law interaction potential correspond to the arrangement of curves from top to bottom near $\Omega = 0$; $\beta = 3$, $\gamma/kv_T = 10^{-2}$; (a) $\nu_{tr}/kv_T = 0.1$ and $n = 1, 2, 4, \infty$; (b) $\nu_{tr}/kv_T = 0.5$ and $n = 1, 2, 3, 4, \infty$; and (c) $\nu_{tr}/kv_T = 3$ and $n = 1, 2, 4, \infty$.

where

$$\nu_{trv}(0) = (1+\beta)\sqrt{\beta}\nu_{eff}, \tag{28}$$

$$\nu_{eff} = \frac{16\sqrt{\pi}LN_b(qq_b)^2}{3M^2v_T^3}, \tag{29}$$

ν_{eff} is the effective transport ion-ion collision frequency, L is the Coulomb logarithm, and q and q_b are the charges of the absorbing and buffer ions in plasma.

Expression (27) for the transport collision frequency can also be derived for the Landau collision integral. To this end, we make use of the well-known [19–22] formulas for the drag force exerted on a test charged particle moving in a medium of charged particles with the Maxwell velocity distribution that correspond to the Landau collision integral. Given the standard [19–22] formulas for **F**, expression (27) for $\nu_{trv}(v)$ follows from relation (15), which defines the transport collision frequency. Thus, the Landau collision integral and the collision integral with the inclusion of pair collisions alone at $n = 1$ give the same expression (27) for the transport collision frequency.

In [9, 10], the line profile was analyzed as a function of the effective transport collision frequency ν_{eff}, which is related to $\nu_{trv}(v)$ at $v = 0$ by $\nu_{eff} = \nu_{trv}(0)/2$ for $\beta = 1$ considered in [9, 10].

In Fig. 4, the central line intensities $I(0)$ (at $\Omega = 0$) calculated in various collision models are plotted against $\nu_{trv}(0)$. As we see from Fig. 4, the difference between the values of $I(0)$ calculated using the Landau integral at $\beta = 1$ and in the kangaroo model at $n = 1$ and $\beta = 1$ (the difference between the two lower solid and dashed curves) is much smaller than the correction that is introduced by allowance for the velocity dependence of the collision frequency (when passing from the strong-collision model to the kangaroo model at $n = 1$ and $\beta = 1$, this is the difference between the two solid curves; when passing from the weak-collision model to the Landau collision integral at $\beta = 1$, this is the differ-

ence between the two dashed curves). The difference between the models is at a maximum when $\gamma \longrightarrow 0$. Therefore, our comparison was made for a small γ/kv_T ($\gamma/kv_T = 10^{-6}$).

Thus, the kangaroo collision model satisfactorily describes the Dicke absorption-line narrowing for ions in equilibrium plasma. Physically, this interesting conclusion follows from the fact that, with a logarithmic accuracy ($\sim 1/L \ll 1$), all collisions in Coulomb plasma up to the impact parameter $\rho = R_D$ may be considered as pair ones, while the contribution of collisions with $\rho > R_D$ may be completely ignored [19, 23]. The collision integral with multiparticle interactions has such a structure that multiple collisions may be essentially considered as pair ones but complicated by the influence of the medium, i.e., by the presence of other particles [19]. In other words, taking into account the combined effect of many perturbing particles on a test particle does not change the binary nature of the scattering formulas; i.e., multiple collisions, on average, imitate pair collisions [22, 24]. In equilibrium plasma, the formulas for multiple collisions match the formulas for pair collisions with a logarithmic accuracy [19, 23]. When calculating the Dicke absorption-line narrowing for ions in equilibrium plasma, the correction due to the difference between the strong- and weak-collision models (the kangaroo model and the Landau collision integral) is much smaller than the correction that is introduced by allowance for the velocity dependence of the collision frequency.

There are several methods for deriving the Landau collision integral. In particular, the Landau collision integral can be derived directly from the Boltzmann kinetic equation (which includes the pair collisions between particles alone), given that collisions with a small change in velocity produces a major effect during Coulomb particle collisions [25, 26]. That is why the above equality of the transport collision frequencies calculated using both formula (27) (pair collisions) and the standard [19–22] expressions for the drag force **F** of a test charged particle in Coulomb plasma comes as no surprise.

The above quantitative comparison of the intensities $I(0)$ calculated in various collision models (see Fig. 4) confirms the possibility of using, to a first approximation, the concept of pair collisions to describe the Dicke absorption-line narrowing for ions in equilibrium plasma.

4. THE ONE-DIMENSIONAL APPROACH TO CALCULATING THE LINE PROFILE

It is well known that when interacting with a gas of resonance particles, laser radiation produces (through the Doppler effect) nonequilibrium in the particle velocity distribution only for the v_z component of particle velocity **v** along the wave vector **k**. The particle distribution in the velocity component $\mathbf{v}_\perp$ orthogonal to

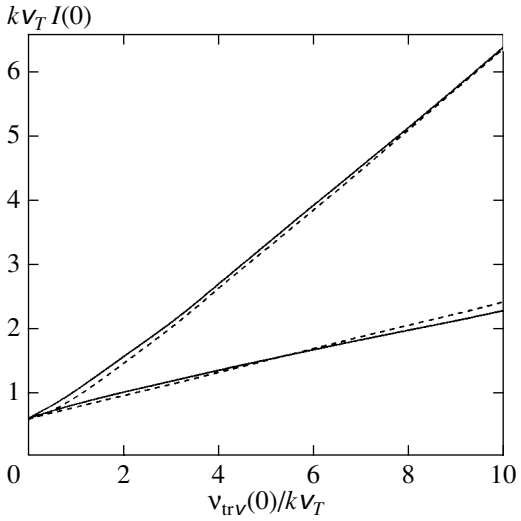

Fig. 4. Central line intensity (for $\Omega = 0$) versus $\nu_{trv}(0)$ in various collision models, $\gamma/kv_T = 10^{-6}$. The upper and lower solid curves represent, respectively, the strong-collision model [the kangaroo model for $\nu_{trv}(v) = \text{const}$] and the kangaroo model, $n = 1$, $\beta = 1$. The upper and lower dashed curves represent, respectively, the weak-collision model (calculations with formulas (2.22) and (2.23) from [2]) and the Landau collision integral (calculations with the interpolation formula (50) from [10]), $\beta = 1$.

the wave vector **k** is not directly perturbed by the radiation. Therefore, in the absence of collisions, the density matrix elements $\rho_{ij}(\mathbf{v})$ for resonance particles can be represented in factorized form:

$$\rho_{ij}(\mathbf{v}) = W(\mathbf{v}_\perp)\rho_{ij}(v_z), \tag{30}$$

where $W(\mathbf{v}_\perp)$ is the Maxwell distribution in velocity component $\mathbf{v}_\perp$. The factorization relation (30) allows the three-dimensional (3D) quantum kinetic equations for the density matrix $\rho_{ij}(\mathbf{v})$ to be reduced to one-dimensional (1D) equations for $\rho_{ij}(v_z)$ by substituting (30) in the basic 3D equations and then integrating them over $\mathbf{v}_\perp$.

When collisions are taken into account, the factorization relation (30) becomes approximate, because collisions generally "transfer" nonequilibrium in the v_z distribution to the distribution in orthogonal velocity component $\mathbf{v}_\perp$. In collision models with velocity-independent collision frequencies, there is no transfer of nonequilibrium and the factorization relation remains valid. Therefore, just as in the absence of collisions, the 3D equations for $\rho_{ij}(\mathbf{v})$ reduce to 1D equations for $\rho_{ij}(v_z)$, which makes it much easier to solve the problem. However, in collision models with velocity-dependent collision frequencies, there is always the transfer of nonequilibrium to the orthogonal velocity components $\mathbf{v}_\perp$ [13]. That is why the factorization relation (30) does not hold in these models, and the equations to be solved must necessarily be 3D ones. In many cases, it is

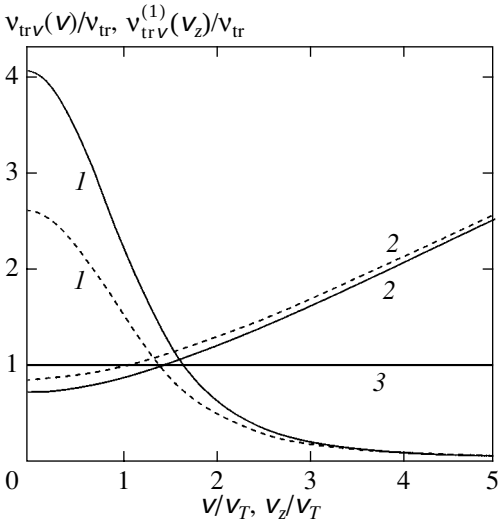

Fig. 5. 3D, $\nu_{trv}(v)$ (solid curves) and 1D, $\nu^{(1)}_{trv}(v_z)$ (dashed curves) transport collision frequencies versus v and v_z; $\beta = 1$; $n = 1$ (*1*), ∞ (*2*), and 4 (*3*).

naturally much more difficult to solve the 3D equations and to analyze their solution than in the 1D situation. Therefore, the question of whether simpler approximate 1D equations can be used to solve spectroscopic and light-induced gas-kinetic problems, in which it is important to take into account the velocity dependence of the collision frequency, becomes of current interest. In other words, it is important to know the error of the solution when passing from the 3D to 1D equations.

It is commonly assumed on intuitive, qualitative grounds that the effect of nonequilibrium transfer is weak [3, 14, 27], and, in many cases, it may be disregarded. In that case, substituting (30) in the basic 3D equations for the density matrix $\rho_{ij}(\mathbf{v})$ and then integrating them over $\mathbf{v}_\perp$ allows us to pass to 1D equations for $\rho_{ij}(v_z)$ with 1D collision integrals:

$$S_{ij}(v_z) = \int S_{ij}(\mathbf{v})\,d\mathbf{v}_\perp, \tag{31}$$

where $S_{ij}(\mathbf{v})$ are the 3D collision integrals.

The accuracy of the solution given by the 1D quantum kinetic equations for velocity-dependent collision frequencies was first quantitatively analyzed in [13, 28]. In these studies, it was numerically shown that the transfer effect is weak and that the transfer of nonequilibrium may be ignored when solving a wide range of spectroscopic and light-induced gas-kinetic problems in which it is important to take into account the velocity dependence of the collision frequency without the risk of losing important subtle details of the described phenomena.

The collisional transfer of nonequilibrium when the phase memory is preserved during collisions has not yet been analyzed quantitatively. In this section, we study this issue using the Dicke effect as an example by directly comparing the numerical calculations of the line profile based on the 3D and 1D equations.

Using the above scheme for passing from the 3D to 1D equations, we find that in the 1D approach to solving our problem, the line profile $I^{(1)}(\Omega)$ is given by formula (10) with the substitution

$$Y_q \longrightarrow Y_q^{(1)}, \quad \langle\tilde{\nu}\rangle \longrightarrow \langle\tilde{\nu}^{(1)}\rangle,$$

where

$$Y_q^{(1)} = \int_{-\infty}^{\infty} \frac{k v_T W(v_z)}{\gamma + \nu^{(1)}(v_z) - i(\Omega - k v_z)} \times \left(\frac{\tilde{\nu}^{(1)}(v_z)}{\langle\tilde{\nu}^{(1)}\rangle}\right)^q dv_z, \tag{32}$$

$$\langle\tilde{\nu}^{(1)}\rangle = \int_{-\infty}^{\infty} \tilde{\nu}^{(1)}(v_z) W(v_z)\,dv_z,$$

$W(v_z)$ is the Maxwell distribution in the velocity component $v_z = \mathbf{k}\cdot\mathbf{v}/k$; $\nu^{(1)}(v_z)$ and $\langle\tilde{\nu}^{(1)}\rangle$ are the 1D collision frequencies. The frequency $\nu^{(1)}(v_z)$ is related to the 3D frequency $\nu(v)$ by

$$\nu^{(1)}(v_z) = \int \nu(v) W(\mathbf{v}_\perp)\,d\mathbf{v}_\perp, \tag{33}$$

the relation between $\tilde{\nu}^{(1)}(v_z)$ and $\tilde{\nu}(v)$ as well as between $\nu^{(1)}_{trv}(v_z)$ and $\nu_{trv}(v)$ are given by similar formulas.

Given the relation

$$\mathbf{v}^2 = \mathbf{v}_\perp^2 + v_z^2$$

formula (33) takes the form

$$\nu^{(1)}(v_z) = \frac{2}{v_T^2}\exp\left(\frac{v_z^2}{v_T^2}\right) \times \int_{|v_z|}^{\infty} v \exp\left(-\frac{v^2}{v_T^2}\right)\nu(v)\,dv. \tag{34}$$

The quantity $\langle\tilde{\nu}^{(1)}\rangle$ in (32) matches $\langle\tilde{\nu}\rangle$ in (11):

$$\langle\tilde{\nu}^{(1)}\rangle = \langle\tilde{\nu}\rangle.$$

It can be shown that if the collision frequencies $\nu(v)$ and $\tilde{\nu}(v)$ are velocity-independent, then the profiles $I(\Omega)$ and $I^{(1)}(\Omega)$ coincide, as must be the case.

Figure 5 shows the v and v_z dependences of the 3D $[\nu_{trv}(v)]$ and 1D $[\nu^{(1)}_{trv}(v_z)]$ transport collision frequencies for a power-law interaction potential at various n. The velocity dependence for the 1D frequency $\nu^{(1)}_{trv}(v_z)$

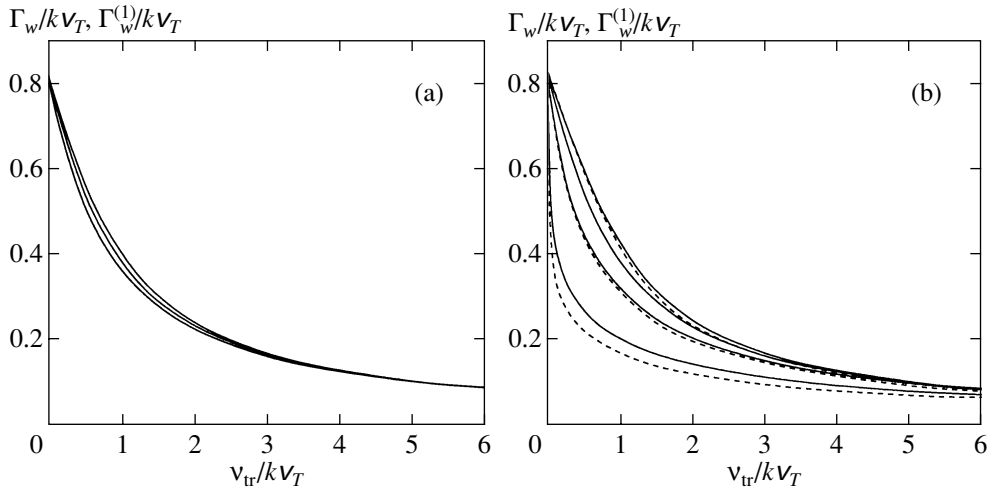

Fig. 6. Line half-widths Γ_w (solid curves) and $\Gamma_w^{(1)}$ (dashed curves; calculations using the 1D equations) versus mean transport collision frequency ν_{tr}. The values of n for a power-law interaction potential correspond to the arrangement of solid curves from top to bottom. $\beta = 10$, $\gamma/kv_T = 10^{-2}$; (a) $n = 6, 4, 3$; (b) $n = \infty, 4, 2, 1$. The plots for Γ_w and $\Gamma_w^{(1)}$ in Fig. 6a merge together and are perceived as a single curve.

is smoother than that for the 3D frequency $\nu_{trv}(v)$. The 1D transport collision frequency decreases with increasing v_z for $n < 4$ and increases with v for $n > 4$. At $n = 4$, the 1D frequency is velocity-independent.

Figure 6 shows the plots of line half-widths Γ_w for the profile $I(\Omega)$ (3D approach) and $\Gamma_w^{(1)}$ for the profile $I^{(1)}(\Omega)$ (1D approach) against mean transport collision frequency ν_{tr} calculated using the above formulas. The calculations were performed under condition (14) [the phase memory was preserved during collisions, $\tilde{\nu}(v) = \nu(v) = \nu_{trv}(v) = \text{const}$)] and for heavy buffer particles, $\beta \gg 1$, because the dependence of the transport collision frequency $\nu_{trv}(v)$ on velocity v is at a maximum in this case, and, hence, the difference between Γ_w and $\Gamma_w^{(1)}$ is also at a maximum. At smaller β, the difference between Γ_w and $\Gamma_w^{(1)}$ is also smaller.

For the power-law potentials with $n = 3$ and $n = 6$, the difference between Γ_w and $\Gamma_w^{(1)}$ is so small ($\lesssim 0.3\%$) that, as we see from Fig. 6a, the plots for Γ_w and $\Gamma_w^{(1)}$ merge together and are perceived as a single curve. The difference between Γ_w and $\Gamma_w^{(1)}$ is marked only for $n \gtrsim 10$ and $n < 3$ (Fig. 6b). However, in these cases, the relative difference between Γ_w and $\Gamma_w^{(1)}$ is also much smaller than the correction that is introduced by allowance for the velocity dependence of the collision frequency (the difference between Γ_w for $n = 4$ and $n \neq 4$).

Thus, when the phase memory is preserved during collisions, the effect of collisional nonequilibrium transfer is weak, and it may be ignored to a first approximation. This implies that the 1D collision integrals can always be used instead of the 3D ones to describe the Dicke narrowing with allowance for the velocity dependence of the collision frequency.

Let us now use the 1D approach to interpret the results obtained in the preceding section. More specifically, let us elucidate the physical cause of the increase in line width with n for a power-law interaction potential. Since the area under the line profile $I^{(1)}(\Omega)$ is constant, an increase in n must cause $I^{(1)}(0)$ to decrease and vice versa. Let us elucidate the cause of this behavior of $I^{(1)}(0)$ as a function of n. Consider a large Doppler absorption-line broadening, $kv_T \gg \nu_{tr}, \gamma$. Under these conditions, only those particles that lie in a narrow velocity range near $v_z = 0$ interact with the field at $\Omega = 0$. Consequently, the diffusion motion of the selected group of particles is determined by the 1D transport collision frequency $\nu_{trv}^{(1)}(0)$. For power-law potentials with $n < 4$, the frequency $\nu_{trv}^{(1)}(0) > \nu_{tr}$ (Fig. 5) and increases with decreasing n [$\nu_{trv}^{(1)}(0) = \nu_{tr}$ for the $n = 4$ potential]. The higher the collision frequency $\nu_{trv}^{(1)}(0)$ of the group of resonance particles near $v_z = 0$, the larger the Dicke line narrowing. Thus, for $n < 4$, the line narrows compared to $n = 4$ [constant transport collision frequency $\nu_{trv}(v) = \text{const}$]. The narrowing increases with decreasing n. The line broadening for $n > 4$ compared to $n = 4$ is explained similarly.

5. CONCLUSION

Our results show that when calculating the Dicke line narrowing, allowance for the velocity dependence of the transport collision frequency, $\nu_{trv}(v)$, can significantly change the calculated line width. The line narrowing or broadening compared to constant collision frequencies can reach 5–12% for molecules, while for ions in equilibrium plasma, the line can narrow by more than half. If $\nu_{trv}(v)$ increases with velocity v, then the line broadens compared to $\nu_{trv}(v)$ = const and narrows if $\nu_{trv}(v)$ decreases with increasing v.

The fact that at some detuning Ω_c of the radiation frequency Ω, $I(\Omega_c)$ was virtually the same for any v dependence of ν_{trv} proved to be unexpected.

Our quantitative calculations have confirmed that simpler approximate 1D quantum kinetic equations can be used instead of the 3D ones to solve spectroscopic problems in which it is important to take into account the velocity dependence of the collision frequency. A quantitative justification of the applicability of the 1D approach to solving such problems is of importance, because the passage from the 3D to 1D equations can give a considerable gain when performing numerical calculations, as recently demonstrated in [11].

ACKNOWLEDGMENTS

We wish to thank S.G. Rautian and D.A. Shapiro for valuable remarks and discussion. This study was supported by the Russian Foundation for Basic Research (project no. 01-02-17433) and the Program of the Ministry of Science of Russia "Physics of Quantum and Wave Processes."

APPENDIX

Let us introduce the Green function for Eq. (2) that satisfies the following equation containing only the parameters of the medium (but not radiation):

$$\gamma F(\mathbf{v}|\mathbf{v}') = S[F(\mathbf{v}|\mathbf{v}')] + \delta(\mathbf{v}-\mathbf{v}'). \quad (A.1)$$

Formally, the solution of Eq. (2) in terms of this Green function is given by

$$\rho(\mathbf{v}) = \int F(\mathbf{v}|\mathbf{v}') \times [W(\mathbf{v}') + i(\Omega - \mathbf{k}\cdot\mathbf{v}')\rho(\mathbf{v}')]d\mathbf{v}'. \quad (A.2)$$

We derived a new integral equation for $\rho(\mathbf{v}')$ that is useful when the velocity distribution $\rho(\mathbf{v}')$ differs only slightly from the Maxwell distribution. Let us show this. Integrating Eq. (2) over the velocities under condition (14) yields

$$(\gamma - i\Omega)\langle\rho\rangle = 1 - i\mathbf{k}\cdot\mathbf{j}, \quad \mathbf{j} \equiv \int \mathbf{v}\rho(\mathbf{v})d\mathbf{v}. \quad (A.3)$$

We calculate $\mathbf{k}\cdot\mathbf{j}$ by using (A.2):

$$\mathbf{k}\cdot\mathbf{j} = \int (\mathbf{k}\cdot\mathbf{v})F(\mathbf{v}|\mathbf{v}') \times [W(\mathbf{v}') + i(\Omega - \mathbf{k}\cdot\mathbf{v}')\rho(\mathbf{v}')]d\mathbf{v}d\mathbf{v}'. \quad (A.4)$$

Integration over $\mathbf{v}$ affects only the Green function:

$$\int \mathbf{v}F(\mathbf{v}|\mathbf{v}')d\mathbf{v} \equiv \mathbf{l}(\mathbf{v}') = \mathbf{v}'\tau(v'). \quad (A.5)$$

The quantities $\mathbf{l}$ and $\tau(v')$ have the dimensions of length and time, respectively (their physical meaning is discussed below). Let us represent $\rho(\mathbf{v})$ as

$$\rho(\mathbf{v}) = \langle\rho\rangle W(\mathbf{v}) + \delta\rho(\mathbf{v}), \quad (A.6)$$

where we assume $\delta\rho(\mathbf{v})$ to be a small addition to the equilibrium distribution $\langle\rho\rangle W(\mathbf{v})$ and disregard this addition when substituting in (A.4). From obvious symmetry considerations, only one term remains when integrating in (A.4) over nonzero $\mathbf{v}'$, so

$$\mathbf{k}\cdot\mathbf{j} = -i\langle\rho\rangle\int(\mathbf{k}\cdot\mathbf{v}')^2\tau(v')W(\mathbf{v}')d\mathbf{v}' = -ik^2D\langle\rho\rangle,$$
$$D \equiv \int(\mathbf{n}\cdot\mathbf{v}')^2\tau(v')W(\mathbf{v}')d\mathbf{v}', \quad (A.7)$$
$$\tau(v') \equiv \int\frac{(\mathbf{v}\cdot\mathbf{v}')}{v'^2}F(\mathbf{v}|\mathbf{v}')d\mathbf{v}.$$

Here, $\mathbf{n}$ is a unit vector in an arbitrarily chosen direction.

Using (A.7), we obtain from (A.3)

$$\langle\rho\rangle = \frac{1}{\gamma + k^2D - i\Omega},$$
$$I(\Omega) \equiv \frac{\mathrm{Re}\langle\rho\rangle}{\pi} = \frac{1}{\pi}\frac{\gamma+\gamma_{tr}}{(\gamma+\gamma_{tr})^2+\Omega^2}, \quad (A.8)$$
$$\gamma_{tr} = k^2D.$$

Thus, at sufficiently high buffer-gas pressures, when the Doppler broadening is negligible and the velocity distribution function $\rho(\mathbf{v})$ is close to the Maxwell one, the line profile $I(\Omega)$ takes a universal form according to (A.8) with the half-width determined by the integrated parameter D. The latter, in turn, is uniquely expressed in terms of $\tau(v)$. The quantities $l(\mathbf{v})$, $\tau(v)$, and D may be considered as phenomenological parameters of the gas system in question, but, as we see from their structure, they are formed by a sequence of collisions, and there is no general expression in terms of the parameters of an elementary scattering event for them. In some particular models, however, they can be expressed in terms of the parameters of the collision integral. Let us consider a model collision integral with a velocity-isotropic arrival:

$$S[\rho(\mathbf{v})] = -\nu(v)\rho(\mathbf{v}) + \int A(v|\mathbf{v}_1)\rho(\mathbf{v}_1)d\mathbf{v}_1, \quad (A.9)$$

where the kernel of the integral is a function of the velocity $\mathbf{v}_1$ before collision and of the velocity magni-

tude $\nu = |\mathbf{v}|$ after collision. The well-known strong-collision model and the kangaroo model considered here are special cases of this model. With this collision integral, we obtain from (A.1) for the Green function

$$F(\mathbf{v}|\mathbf{v}') = \frac{1}{\gamma + \nu(\nu)} \times [\delta(\mathbf{v}-\mathbf{v}') + \int A(\nu|\mathbf{v}_1)F(\mathbf{v}_1|\mathbf{v}')d\mathbf{v}_1]. \quad \text{(A.10)}$$

The Green function consists of two parts: anisotropic (with a δ function) and isotropic in $\mathbf{v}$ (integral term). The anisotropic part is completely determined by the model parameters. On the other hand, $\tau(\nu')$ from (A.7) is entirely determined by the anisotropic part of the Green function alone, so, based on (A.10), we derive for model (A.9)

$$\tau(\nu') = \frac{1}{\gamma + \nu(\nu')}. \quad \text{(A.11)}$$

Note that $\tau(\nu')$ has also a finite limit when $\gamma \longrightarrow 0$, whereas $\int F(\mathbf{v}|\mathbf{v}')d\mathbf{v}$ tends to infinity in this limit.

Let us now turn to the physical meaning of l, D, and τ. To this end, we consider the standard Boltzmann kinetic equation with a stationary source and relaxation:

$$(\gamma + \mathbf{v}\cdot\nabla)\rho(\mathbf{v}) = S[\rho(\mathbf{v})] + Q(\mathbf{r})W(\mathbf{v}), \quad \text{(A.12)}$$

where $\rho(\mathbf{v})$ is coordinate-dependent. We use the Green function (A.1) to solve it. Formally, the solution can be written as

$$\rho(\mathbf{v}) = \int F(\mathbf{v}|\mathbf{v}') \times [Q(\mathbf{r})W(\mathbf{v}') - (\mathbf{v}'\cdot\nabla)\rho(\mathbf{v}')]d\mathbf{v}'. \quad \text{(A.13)}$$

Let us calculate the particle flux

$$\mathbf{j} \equiv \int \mathbf{v}\rho(\mathbf{v})d\mathbf{v},$$

by assuming the local deviation from the Maxwell distribution to be negligible. Acting in the same way as in the case of deriving (A.4)–(A.7), we obtain

$$\mathbf{j} = -\int l(\mathbf{v}')[(\mathbf{v}'\cdot\nabla)\langle\rho\rangle]W(\mathbf{v}')d\mathbf{v}' = -(\int \tau(\nu')(\mathbf{n}\cdot\mathbf{v}')^2 W(\mathbf{v}')d\mathbf{v}')\nabla\langle\rho\rangle = -D\nabla\langle\rho\rangle. \quad \text{(A.14)}$$

It is clear from this result that D is the ordinary diffusivity when $\gamma \longrightarrow 0$, which is defined as the coefficient of proportionality between the density gradient and the particle flux produced by it. The diffusivity can be expressed in terms of the transport parameters averaged in a certain way—transport length $l(\mathbf{v}')$ or transport time $\tau(\nu')$. The meaning of the transport length can be understood from its definition (A.5). In accordance with the normalization of the Green function given by Eq. (A.1), the combination $F(\mathbf{v}|\mathbf{v}')d\mathbf{v}$ means the average time a particle with velocity $\mathbf{v}$ lies in interval $d\mathbf{v}$. The combination $\mathbf{v}F(\mathbf{v}|\mathbf{v}')d\mathbf{v}$ is the average particle displacement dl in this time. Consequently, $l(\mathbf{v}')$ is the average displacement of a particle with an initial velocity $\mathbf{v}'$ until the memory of this velocity is completely lost. In general, this requires a finite number of collisions. The meaning of $\tau(\nu')$ is also transparent: this is the average time it takes to lose the memory of the initial velocity ($\mathbf{v}'$) direction. The transport parameters $l(\mathbf{v}')$ and $\tau(\nu')$ were previously introduced (see, e.g., [29, 30]) in connection with gas-kinetic problems. They are seen to have a direct bearing on the line-profile problem as well.

REFERENCES

1. R. H. Dicke, Phys. Rev. **89**, 472 (1953).
2. S. G. Rautian and I. I. Sobel'man, Usp. Fiz. Nauk **90**, 209 (1966) [Sov. Phys. Usp. **9**, 701 (1966)].
3. S. G. Rautian, G. I. Smirnov, and A. M. Shalagin, *Nonlinear Resonances in Atomic and Molecular Spectra* (Nauka, Novosibirsk, 1979).
4. L. Galatry, Phys. Rev. **122**, 1218 (1961).
5. S. G. Rautian, Opt. Spektrosk. **86**, 385 (1999) [Opt. Spectrosc. **86**, 334 (1999)].
6. A. Henry, D. Hurtmans, M. Margottin-Maclou, and A. Valentin, J. Quant. Spectrosc. Radiat. Transf. **56**, 647 (1996).
7. B. Lance, G. Blanquet, J. Walrand, and J. P. Bouanich, J. Mol. Spectrosc. **185**, 262 (1997).
8. A. S. Pine, J. Quant. Spectrosc. Radiat. Transf. **62**, 397 (1999).
9. E. V. Podivilov and D. A. Shapiro, Pis'ma Zh. Éksp. Teor. Fiz. **56**, 465 (1992) [JETP Lett. **56**, 449 (1992)].
10. E. V. Podivilov, A. I. Chernykh, and D. A. Shapiro, Zh. Éksp. Teor. Fiz. **105**, 1214 (1994) [JETP **78**, 653 (1994)].
11. D. A. Shapiro and A. D. May, Phys. Rev. A **63**, 012701 (2001).
12. A. Brissaud and U. Frish, J. Math. Phys. **15**, 524 (1974).
13. T. Privalov and A. Shalagin, Phys. Rev. A **59**, 4331 (1999).
14. S. G. Rautian and A. M. Shalagin, *Kinetic Problems of Nonlinear Spectroscopy* (North-Holland, Amsterdam, 1991).
15. F. Kh. Gel'mukhanov, L. V. Il'ichov, and A. M. Shalagin, Physica A (Amsterdam) **137**, 502 (1986).
16. E. M. Lifshitz and L. P. Pitaevskiĭ, *Physical Kinetics* (Nauka, Moscow, 1979; Pergamon, Oxford, 1981).
17. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 1: *Mechanics* (Nauka, Moscow, 1988; Pergamon, New York, 1988).
18. V. E. Golant, A. P. Zhilinskiĭ, and I. E. Sakharov, *Fundamentals of Plasma Physics* (Atomizdat, Moscow, 1977; Wiley, New York, 1980).
19. B. A. Trubnikov, in *Reviews of Plasma Physics*, Ed. by M. A. Leontovich (Gosatomizdat, Moscow, 1963; Consultants Bureau, New York, 1963), issue 1, p. 98.
20. B. A. Trubnikov, *Theory of Plasma* (Énergoatomizdat, Moscow, 1996).

21. S. Ishimaru, *Basic Principles of Plasma Physics* (Benjamin, New York, 1973; Atomizdat, Moscow, 1975).

22. F. Hinton, in *Handbook of Plasma Physics*, Vol. 1: *Basic Plasma Physics*, Ed. by A. A. Galeev and R. N. Sudan (Énergoatomizdat, Moscow, 1983; North-Holland, Amsterdam, 1983), p. 152.

23. G. Ecker, *Theory of Fully Ionized Plasmas* (McGraw-Hill, New York, 1972; Mir, Moscow, 1974).

24. V. I. Kogan, Dokl. Akad. Nauk SSSR **135**, 1374 (1960) [Sov. Phys. Dokl. **5**, 1316 (1960)].

25. V. P. Silin, *An Introduction to the Kinetic Theory of Gases* (Nauka, Moscow, 1971).

26. V. M. Zhdanov, *Transport Phenomena in Multicomponent Plasma* (Énergoizdat, Moscow, 1982).

27. P. R. Berman, T. W. Mossberg, and S. R. Hartmann, Phys. Rev. A **25**, 2550 (1982).

28. A. I. Parkhomenko and A. M. Shalagin, Zh. Éksp. Teor. Fiz. **118**, 279 (2000) [JETP **91**, 245 (2000)].

29. I. Kuščer, L. J. F. Hermans, P. L. Chapovsky, *et al.*, J. Phys. B **26**, 2837 (1993).

30. B. Nagels, P. L. Chapovsky, L. J. F. Hermans, *et al.*, Phys. Rev. A **53**, 4305 (1996).

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 93, No. 4, 2001, pp. 737–752.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 846–862.

ATOMS, SPECTRA, RADIATION

Short Optical Pulse Polarization Dynamics in a Nonlinear Birefringent Doped Fiber¶

S. O. Elyutin* and A. I. Maimistov**
Moscow State Engineering Physics Institute, Moscow, 115409 Russia
**e-mail: elyutin@star.mephi.ru, **sergeipe@mtu-net.ru*
Received May 29, 2001

Abstract—Numerical solutions are obtained of the full self-consistent system of equations for the counter, rotating polarization components of the field of a short optical pulse propagating in a nonlinear birefringent fiber and in the ensemble of the energy-level degenerate doped resonance atoms implanted in the fiber material. In every cross section of the fiber, the ellipticity of the polarized wave experiences a complex evolution in time accompanied by rapid changes of the azimuthal angle due to the interplay of the dispersion and the Kerr nonlinear self- and cross-phase modulation. The reciprocal effect of the impurities on the traveling pulse causes oscillations of the pulse envelope that can completely distort the shape of the input signal, while the resonance absorption can drive the birefringence process from the nonlinear regime back to the linear one. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Much is known about the propagation of short optical pulses in nonlinear fibers [1–3]. In a nonlinear optical fiber, the propagation of distortionless pulses can be realized under conditions where the amplitude self-modulation effects compensate for the linear dispersion. In particular, for intensities at which the dielectric polarizability has a cubic field response (the Kerr nonlinearity), the envelopes of quasimonochromatic pulses are approximated by optical solitons. In the axisymmetric optical fiber, the fundamental mode consists of two copropagating and perpendicularly polarized linear fields. Non-axisymmetric imperfections to the fiber destroy this polarization degeneracy and introduce the linear birefringence—a difference in the propagation characteristic between the two polarizations. Furthermore, for a nonlinear fiber, the amplitude coupling causes an additional self-induced birefringence via the cross-phase modulation. Activating the fiber by resonance impurities, e.g., rare-earth ions, has given rise to an entire industry of fiber lasers and amplifiers whose physics is extensively discussed in the literature (see [4] and references therein).

In this paper, our approach is to consider the physical system in two associated parts. The first is a short optical pulse propagating in a nonlinear, dispersive and birefringent fiber. The second is an ensemble of two-level resonance atoms immersed in the fiber host material. The first part is modeled by the full fiber equations, broadly in the form of two coupled nonlinear Schrödinger (NLS) equations. The second is governed by a system of Bloch equations coupled to the fiber part by the resonance polarization.

Disregarding the linear birefringence, the difference of the group velocities of the polarized modes (i.e., the walk-off effect), and the polarization induced in the resonant subsystem, the nonlinear equations for the field components are an example of a completely integrable system [5, 6]. Under certain conditions, a short optical pulse in a resonant medium can in turn evolve into a steady-state solitary wave (a 2π-pulse) [7]. This means that in a model of this type, one could ideally observe the coexistence of the self-induced transparency (SIT) and NLS solitons [8–10]. But this can hardly occur for a moderately intense pulse in realistic doped fibers because the disparity in the spatial scales and the pulse energy for the SIT solitons and optical solitons are very substantial in a nonlinear fiber: one 2π SIT pulse corresponds to hundreds of NLS solitons by power.

The general problem then apparently reduces to two characteristic cases: (i) the Kerr nonlinearity and birefringence effect on the coherent propagation of short pulses in a short doped fiber when the dispersion of group velocities and the walk-off effect are insignificant, (ii) the weak effect of the resonant absorption and refraction on the coupled soliton-like pulse propagation in a nonlinear birefringent fiber. In this paper, we concentrate on the latter case. Basically, both the fiber effects and the two-level medium can affect the polarization state of the propagating light wave. In vector nonlinear wave equations, all the cross terms (linear birefringence, power-dependent cross-phase modulation, and the group velocity mismatch) contribute to variations of the polarization states [11]. At the same time, level-degenerate atoms possess their own polarization properties that have been discussed in numerous papers devoted, e.g., to polarization features of the photon–echo effect [2]. The eccentricity and polarization

¶This article was submitted by the authors in English.

ellipse orientation can alter across the width of a 2π-pulse in degenerate self-induced transparency [12, 13].

In this paper, we consider a short optical pulse propagation in a nonlinear birefringent doped fiber by numerically solving the self-consistent system of equations for the optical field and the degenerate two-level medium. In the course of discussion, we introduce in succession the conventional fiber attributes (birefringence, dispersion, Kerr nonlinearity, and walk-off effect) followed by the resonance interaction of the light pulse with the impurity atoms in order to observe both the separate and the combined influence of these effects on the dynamics of polarization states and the waveforms of polarization modes.

2. POLARIZED WAVES IN A CUBIC MEDIUM WITH RESONANT IMPURITIES

We consider the electromagnetic wave propagation in an optical birefringent fiber with the cubic (Kerr) nonlinearity. We let this fiber contain doped two-level atoms with the transition energy in resonance with the frequency of the carrier. The description of the solitary wave propagation is conventionally based on the reduced Maxwell equations [9, 14–16] complemented with the Bloch equations [5] determining the evolution of the resonance subsystem. Hereafter, we follow the works by Boardman and Cooper [9, 14], where the propagation of polarized pulses in the Kerr medium was thoroughly observed. The resonance contribution is considered in the same way as in [5].

We write the electric field vector of the optical wave as $\mathbf{E} = E_x\mathbf{e}_x + E_y\mathbf{e}_y$, where $\mathbf{e}_x$ and $\mathbf{e}_y$ are orthogonal vectors in the x and y directions. The wave propagates in the z direction. Using the slowly varying (complex) envelope approximation (SVEA), we can write

$$E_x = \mathscr{E}_x(z, t)\Psi(x, y)\exp[i(\beta_x z - \omega_0 t)],$$
$$E_y = \mathscr{E}_y(z, t)\Psi(x, y)\exp[i(\beta_y z - \omega_0 t)],$$

where ω_0 is the carrier frequency and $\beta_x(\beta_y)$ is the linear propagation constant of the slow (fast) principal axis of the birefringent fiber [9, 14–16].

The radial distribution of the electric field in the fiber is described by the mode function $\Psi(x, y)$. We assume that the propagation constants slightly vary from some average value β such that $\beta_x = \beta + \Delta\beta$ and $\beta_y = \beta - \Delta\beta$. The complex envelopes are expressed in terms of real amplitudes and phases slowly varying in space and time, $\mathscr{E}_x = R_x\exp[i\varphi_x]$ and $\mathscr{E}_y = R_y\exp[i\varphi_y]$. The phases are given by $\varphi_x = \tilde{\varphi} + \phi$ and $\varphi_y = \tilde{\varphi} - \phi$, where $\tilde{\varphi}$ is the average value. Finally, the electric field components are given by

$$E_x = A_x(z, t)\Psi(x, y)\exp[i\beta z - i\omega_0 t],$$
$$E_y = A_y(z, t)\Psi(x, y)\exp[i\beta z - i\omega_0 t],$$

where

$$A_x = R_x\exp[i(\tilde{\varphi} + \phi + \Delta\beta z)],$$
$$A_y = R_y\exp[i(\tilde{\varphi} - \phi - \Delta\beta z)].$$

The presentation of the phase terms in the above form is attributable to the following effects: the intrinsic birefringence (i.e., the birefringence that would exist in the linear limit) is represented by $\pm\Delta\beta z$ and the self-phase modulation is described by $\pm\phi(t, z)$. In a weakly nonlinear and weakly birefringent medium, such as the typical glasses of optical fibers, the nonlinearity is assumed to be instantaneous. Generally speaking, the validity of this assumption depends on the pulse rise time. If the optical pulse becomes narrower, the assumption is no longer valid. We ignore this effect in this paper. We consider a Kerr-type nonlinear medium assuming that (i) the dielectric medium is isotropic, (ii) the third harmonic generation can be neglected, and (iii) the second-order nonlinear susceptibility is identically zero. Therefore, the slowly varying envelope of the nonresonance cubic polarization $\mathscr{P}^{\text{Kerr}}(z, t)$ is

$$\mathscr{P}^{\text{Kerr}} = 2\chi^{(3)}_{1122}(\mathscr{E}\cdot\mathscr{E}^*) + \chi^{(3)}_{1221}(\mathscr{E}\cdot\mathscr{E})\mathscr{E}^*,$$

where the spatial and temporal dispersion is assumed to be absent. The above equality written in projections becomes

$$P_x^{\text{Kerr}} = [(a + b)|A_x|^2 + \{a + b\exp[-4i(\phi + \Delta\beta z)]\}|A_y|^2]A_x, \quad (1)$$

$$P_y^{\text{Kerr}} = [(a + b)|A_y|^2 + \{a + b\exp[4i(\phi + \Delta\beta z)]\}|A_x|^2]A_y, \quad (2)$$

where we use the notation $2\chi^{(3)}_{1122} = a$ and $\chi^{(3)}_{1122} = b$.

The Maxwell equations with nonlinear terms (1) and (2) and the resonance polarization included provide a set of coupled evolutionary equations for the amplitudes A_x and A_y, with the second-order group velocity dispersion taken into account:

$$i\frac{\partial A_x}{\partial z} + iv_x^{-1}\frac{\partial A_x}{\partial t} - \sigma_x\frac{\partial^2 A_x}{\partial t^2} + \Delta\beta A_x + m_x(a_{11}|A_x|^2 + a_{12}|A_y|^2)A_x + qP_x = 0, \quad (3)$$

$$i\frac{\partial A_y}{\partial z} + iv_y^{-1}\frac{\partial A_y}{\partial t} - \sigma_y\frac{\partial^2 A_y}{\partial t^2} - \Delta\beta A_y + m_y(a_{21}|A_x|^2 + a_{22}|A_y|^2)A_y + qP_y = 0. \quad (4)$$

In Eqs. (3) and (4), the effect of the resonance impurities is referred to by the slowly varying polarization envelopes P_x and P_y. The coefficient is defined in Eq. (15) in what follows. Nonresonance losses are ignored in (3) and (4), while the terms proportional to

P_x and P_y represent the resonance absorption and refraction effects. In Eqs. (3) and (4), the following coefficients are introduced:

$$v_{x,y}^{-1} = \frac{d\beta_{x,y}}{d\omega}, \quad \sigma_{x,y} = \frac{1}{2}\frac{d^2\beta_{x,y}}{d\omega^2}, \quad m_{x,y} = \frac{\omega_0^2}{2c^2\beta_{x,y}}.$$

The self-modulation effect is represented by a_{11} and a_{22}. The factors a_{12} and a_{21} are responsible for the cross-modulation. The effective nonlinear interaction parameter χ_{eff} is taken as a factor exclusively depending on the ratio of the material susceptibility tensor elements, i.e., on $\chi_{1221}^{(3)}/\chi_{1122}^{(3)}$. Thus, we have

$$a_{11} = a_{22} = \chi_{\text{eff}}, \quad a_{12} = \frac{a + b\exp[-4i(\phi + \Delta\beta z)]}{a + b}\chi_{\text{eff}},$$

$$a_{21} = \frac{a + b\exp[4i(\phi + \Delta\beta z)]}{a + b}\chi_{\text{eff}},$$

where the effective nonlinear interaction parameter χ_{eff} is defined as

$$\chi_{\text{eff}} = \frac{\int \chi_{1122}^{(3)}(\boldsymbol{\rho})|\Psi(\boldsymbol{\rho})|^4 d\boldsymbol{\rho}}{\int |\Psi(\boldsymbol{\rho})|^2 d\boldsymbol{\rho}}.$$

For the silica optical fiber, the third-order susceptibility mainly occurs because of a nonlinear electronic response and because $a = 2b$, and therefore,

$$a_{12} = \chi_{\text{eff}}\left(\frac{2}{3} + \frac{1}{3}\exp[-4i(\phi + \Delta\beta z)]\right),$$

$$a_{21} = \chi_{\text{eff}}\left(\frac{2}{3} + \frac{1}{3}\exp[4i(\phi + \Delta\beta z)]\right).$$

If we use the relations

$$\exp[4i(\phi + \Delta\beta z)] = \frac{E_x E_y^*}{E_x^* E_y} = \frac{A_x A_y^*}{A_x^* A_y},$$

$$\exp[-4i(\phi + \Delta\beta z)] = \frac{E_y E_x^*}{E_y^* E_x} = \frac{A_y A_x^*}{A_y^* A_x},$$

the nonlinear terms in Eqs. (3) and (4) become

$$(a_{11}|A_x|^2 + a_{12}|A_y|^2)A_x$$
$$= \chi_{\text{eff}}\left(|A_x|^2 A_x + \frac{2}{3}|A_y|^2 A_x + \frac{1}{3}A_x^* A_y^2\right),$$

$$(a_{21}|A_x|^2 + a_{22}|A_y|^2)A_y$$
$$= \chi_{\text{eff}}\left(|A_y|^2 A_y + \frac{2}{3}|A_x|^2 A_x + \frac{1}{3}A_y^* A_x^2\right).$$

The system of equations (3), (4) can now be rewritten in the final form

$$i\frac{\partial A_x}{\partial z} + iv_x^{-1}\frac{\partial A_x}{\partial t} - \sigma_x\frac{\partial^2 A_x}{\partial t^2} + \Delta\beta A_x + m_x\chi_{\text{eff}}\left(|A_x|^2 A_x + \frac{2}{3}|A_y|^2 A_x + \frac{1}{3}A_x^* A_y^2\right) + qP_x = 0, \tag{5}$$

$$i\frac{\partial A_y}{\partial z} + iv_y^{-1}\frac{\partial A_y}{\partial t} + -\sigma_y\frac{\partial^2 A_y}{\partial t^2} - \Delta\beta A_y + m_y\chi_{\text{eff}}\left(|A_y|^2 A_y + \frac{2}{3}|A_x|^2 A_x + \frac{1}{3}A_y^* A_x^2\right) + qP_y = 0. \tag{6}$$

Equations (5) and (6) describe the propagation of a polarized radiation pulse in the birefringent fiber doped by resonance impurities. The electric field of the pulse is expressed by the Cartesian components. In order to emphasize the circular nature of birefringence, it seems reasonable to express the evolution equations in terms of the right- and left-hand circularly polarized fields

$$E_1 = E_x + iE_y, \quad E_2 = E_x - iE_y,$$

The corresponding complex envelopes can be written as

$$A_1 = A_x + iA_y, \quad A_2 = A_x - iA_y.$$

It is worth noting that

$$A_x^2 + A_y^2 = A_1 A_2, \quad |A_x|^2 + |A_y|^2 = \frac{|A_1|^2 + |A_2|^2}{2}.$$

Hereafter, we set $\sigma_x = \sigma_y = \sigma$ and $m_x = m_y = m$. For the sake of generality, we also assume that the group velocities of the different polarization components are different. To proceed to numerical simulations of the nonlinear propagation of polarized pulses, it is convenient to introduce the dimensionless quantities

$$A_{1,2} = A_0 e_{1,2}, \quad z = \zeta L, \quad \tau = \left(t - \frac{z}{v}\right)t_0^{-1},$$

where t_0 is the characteristic time scale (which can be equal to the initial pulse duration t_{p0}); L and A_0 are the normalizing length and amplitude, respectively; and the velocity of the time frame v is the velocity of the "center of gravity" of the optical pulse, $v^{-1} = (v_1^{-1} + v_2^{-1})/2$.

In terms of the new variables, the system of equations (4) becomes

$$i\frac{\partial e_1}{\partial \zeta} + i\frac{1}{l_g}\frac{\partial e_2}{\partial \tau} - \frac{s}{l_d}\frac{\partial^2 e_1}{\partial \tau^2} + \frac{1}{l_c}e_2 + \frac{1}{3l_k}(|e_1|^2 + 2|e_2|^2)e_1 + \left(\frac{Lq}{A_0}\right)P_1 = 0, \tag{7}$$

$$i\frac{\partial e_2}{\partial \zeta} + i\frac{1}{l_g}\frac{\partial e_1}{\partial \tau} - \frac{s}{l_d}\frac{\partial^2 e_2}{\partial \tau^2} + \frac{1}{l_c}e_1 + \frac{1}{3l_k}(|e_2|^2 + 2|e_1|^2)e_2 + \left(\frac{Lq}{A_0}\right)P_2 = 0, \quad (8)$$

where $P_1 = P_x + iP_y$ and $P_2 = P_x - iP_y$.

In Eqs. (7) and (8), the effect of resonance impurities is represented by the slowly varying polarization envelopes P_1 and P_2. The parameters l_g, l_c, l_k, and l_d are

$$l_g^{-1} = \frac{L}{L_g} = \frac{L}{2t_0}\left(\frac{1}{v_1} - \frac{1}{v_2}\right), \quad l_c^{-1} = \frac{L}{L_c} = \Delta\beta L,$$
$$l_k^{-1} = \frac{L}{L_k} = L\chi_{\text{eff}}A_0^2\frac{\omega_0^2}{2c^2\beta}, \quad (9)$$
$$s = \operatorname{sgn}\sigma, \quad l_d^{-1} = \frac{L}{L_d} = \frac{L}{t_0^2}|\sigma|,$$

where

$$L_d = \frac{t_0^2}{|\sigma|}, \quad L_c = \frac{1}{\Delta\beta},$$
$$L_k = \frac{2\beta c^2}{\omega_0^2\chi_{\text{eff}}A_0^2}, \quad L_g = \frac{2v_1v_2t_0}{v_2 - v_1}. \quad (10)$$

The length L_d characterizes the dispersion of the group velocities in each polarization mode. The quantity L_c stands for the coupling length. The corresponding terms in Eqs. (7) and (8) couple the right and left circular components of the electromagnetic wave, thereby implying the linear birefringence effect. The self- and cross-modulation effects reveal at the length L_k. The difference between the group velocities v_1 and v_2 of the counter-rotating polarized light waves causes a spatial divergence of the differently polarized components of the optical pulse (the walk-off effect) over the characteristic length L_g. A simple estimate gives the ratio

$$\frac{l_c}{l_g} \approx \frac{\lambda}{4\pi(ct_{p0})}.$$

It follows that in the picosecond pulse range, the terms related to the group velocities are small compared to the linear coupling terms. However, the walk-off effect can be important when we pass to the femtosecond pulse duration domain.

Equations (7) and (8) must be supplemented by equations describing the temporal behavior of the density matrix elements for the ensemble of two-level atoms immersed in the fiber host material whose levels are degenerate with respect to the projections of the angular momenta j_a and j_b. For definiteness, we consider the case where $j_a = 1 \longrightarrow j_b = 0$. The resonance transition is characterized by the dipole moment element $d_{13} = d_{23} = d_{31}^* = d_{32}^* = d$. The effective matrix element of the dipole transition is given by

$$d_{\text{eff}} = \frac{\int d(\boldsymbol{\rho})|\Psi(\boldsymbol{\rho})|^2 d\boldsymbol{\rho}}{\int |\Psi(\boldsymbol{\rho})|^2 d\boldsymbol{\rho}}.$$

The vector $\boldsymbol{\rho}$ lies in the plane normal to the optical fiber axis.

The temporal behavior of the resonant impurities is governed by a system of the generalized Bloch equations [13]. For slowly varying elements of the density matrix $\hat{\rho}$ describing the transition between the states $|a, m\rangle = |j_a = 1, m = \pm 1\rangle$, and $|b\rangle = |j_b = 0, m = 0\rangle$, we introduce the notation

$$\rho_{12} = \langle a, -1|\hat{\rho}|a, +1\rangle, \quad \rho_{13} = \langle a, -1|\hat{\rho}|b\rangle,$$
$$\rho_{23} = \langle a, +1|\hat{\rho}|b\rangle, \quad \rho_{11} = \langle a, -1|\hat{\rho}|a, -1\rangle,$$
$$\rho_{22} = \langle a, +1|\hat{\rho}|a, +1\rangle, \quad \rho_{33} = \langle b|\hat{\rho}|b\rangle, \quad (11)$$
$$\rho_{kl} = \rho_{lk}^*, \quad l, k = 1, 2, 3.$$

The initial conditions are given by

$$\rho_{33}(0) = 1, \quad \rho_{22}(0) = \rho_{11}(0) = 0,$$
$$\rho_{12}(0) = \rho_{13}(0) = \rho_{23}(0) = 0.$$

The change of variables $\rho_{12} = m_{21}$, $\rho_{21} = m_{12}$, $\rho_{11} = m_{11}$, $\rho_{22} = m_{22}$, $\rho_{33} = n$, $p_1 = -\rho_{13}$, $p_2 = -\rho_{23}$ allows writing the generalized system of Bloch equations in the compact form

$$\frac{\partial p_\alpha}{\partial \tau} = i\nu p_\alpha - if\left(\sum_{\alpha'} e_{\alpha'}m_{\alpha'\alpha} - e_\alpha n\right), \quad (12)$$

$$\frac{\partial m_{\alpha\alpha'}}{\partial \tau} = -if(e_\alpha^* p_{\alpha'} - e_{\alpha'}p_\alpha^*), \quad (13)$$

$$\frac{\partial n}{\partial \tau} = -if\sum_\alpha (e_\alpha p_\alpha^* - e_\alpha^* p_\alpha), \quad \alpha, \alpha' = 1, 2. \quad (14)$$

The initial conditions are given by $n(0) = 1$ for the ground level population, $p_\alpha(0) = 0$ for the polarization, and $m_{\alpha\alpha'}(0) = 0$. We also assume that the pulse duration is much shorter than all the relaxation times in the resonance subsystem, which allows us to omit the relaxation terms in Eqs. (12)–(14). The dimensionless variables p_α entering Eqs. (12)–(14) are related to the polarization terms in Eqs. (7) and (8) by

$$\left(\frac{Lq}{A_0}\right)P_\alpha = \frac{L}{L_r}P_\alpha = \frac{L}{L_r}\langle p_\alpha\rangle = \frac{1}{l_r}\langle p_\alpha\rangle, \quad (15)$$

where

$$q = \frac{2\pi\omega_0 n_a d_{\text{eff}}}{cn(\omega_0)},$$

$L_r = fL_r^{(2\pi)}$, n_a is the concentration of the impurity atoms, and $\langle\ \rangle$ denotes the summation over all atoms with the frequency detunings $\nu = \Delta\omega t_0$ from the center of the inhomogeneously broadened line. In (15), the characteristic length of the resonance interaction is

$$L_r^{(2\pi)} = \frac{cn\hbar}{\pi d_{\text{eff}}^2 \omega_0 n_a t_0}. \quad (16)$$

In the system of equations (12)–(13) and in expression (15),

$$f = \frac{d_{\text{eff}} A_0 t_0}{2\hbar} = \frac{A_0}{A_{2\pi}}$$

is the normalized effective oscillation frequency of the material variables of the resonance subsystem affected by the field of the amplitude A_0, and $A_{2\pi}$ is the amplitude of the SIT 2π-pulse.

The coupled system of Maxwell–Bloch equations (7)–(8) and (12)–(14) provides the mathematical basis for numerically simulating the propagation of short pulses of circularly polarized light in a nonlinear waveguide doped by resonance impurities. The solution of field equations (7)–(8) was obtained using one of the popular [17] finite difference implicit–explicit Crank–Nicolson numerical schemes, where the desired accuracy 0.001 was reached by iterations. Bloch equations (12)–(14) coupled to field equations (7)–(8) by the resonance polarization terms were solved by the predictor–corrector procedure. The predictor–corrector was run at every iteration in the Crank–Nicolson algorithm until the accuracy about 0.001 was achieved for the polarization components p_α in Eqs. (12)–(14). Although the code could produce the integration over the inhomogeneously broadened line of the resonance absorption, we restricted it to the homogeneous case and the exact resonance at this stage of numerical simulation, i.e., to $\nu = 0$ in (12). The results of calculations were the absolute value of the complex amplitudes $e_{1,2}(\zeta, \tau)$ of the counter-rotating right- and left-handed oppositely polarized fields. Following Winful [18], we examined the polarization state of the field in the optical pulse in terms of the azimuthal angle

$$\theta(\zeta, \tau) = \frac{\arg\xi}{2}$$

and the ellipticity

$$\varepsilon(\zeta, \tau) = \frac{|\xi| - 1}{|\xi| + 1},$$

where $\xi = e_1 e_2^{-1}$ is a complex quantity. The characteristic values of ε are given by $\varepsilon = 0$ for the linearly polarized light, $\varepsilon = +1$ for the purely right-hand circularly polarized light, and $\varepsilon = -1$ for the purely left-hand circularly polarized light. The parameter θ is the angle between the axis of the polarization ellipse and the slow principal axis of the birefringent fiber. It can vary within the interval $(-\pi/4, \pi/4)$.

The launched pulses are assumed to have the following form,

$$e_{1,2}(0, \tau) = e_{m1,2}\,\text{sech}\left(\frac{\tau - \tau_0}{\delta}\right), \quad (17)$$

where $\delta = t_{p0} t_0^{-1}$ and τ_0 is the temporal coordinate of the input pulse center.

3. NUMERICAL ESTIMATES

We let the group velocity dispersion $D = 4\pi c\sigma\lambda_0^{-2}$ of the silica-based monomode fiber host material be typically $D = 15$ ps nm^{-1} km^{-1} at $\lambda_0 = 1.55$ μm and the nonlinear index $n_2 \approx 10^{-13}$ esu. It then follows that

$$\sigma = \frac{1}{2}\left|\frac{d^2\beta}{d\omega^2}\right| \approx 10^{-28}\ \text{s}^2\ \text{cm}^{-1}.$$

The effective nonlinear interaction parameter is

$$\chi_{\text{eff}} \approx \frac{n_2 n}{2\pi} \approx 2.3 \times 10^{-14}\ \text{esu}.$$

We adopt the value $d \approx 5 \times 10^{-21}$ esu (the transition $^4I_{5/2} \longrightarrow {}^4I_{5/2}$ in Er^{3+} ions) and the impurity concentration $n_a \approx 10^{18}$ cm^{-3} that corresponds to realistic samples [19]. With the input pulse duration $t_{p0} = t_0 = 0.1$ ps, the dispersion length is $L_d = t_0^2\sigma^{-1} \approx 10^2$ cm. The polarization mode coupling effect occurs over the distance

$$L_c \approx \frac{1}{\Delta\beta} \approx \frac{\lambda_0}{2\pi\Delta n} \approx 25\ \text{cm}.$$

Here, we set $\Delta n \sim 10^{-6}$ [20]. The effect of the group velocity mismatch becomes noticeable at the characteristic distance

$$L_g = \frac{2v_1 v_2}{v_2 - v_1} t_0 \approx \frac{2ct_0}{\Delta n} \approx 6 \times 10^4\ \text{cm}.$$

The spatial scale of the Kerr self- and cross-modulation process L_k depends on the field amplitude A_0 as

$$L_k \approx \frac{n\lambda_0}{\pi\chi_{\text{eff}} A_0^2}.$$

The balance between the fiber group velocity dispersion and the nonlinear pulse compression occurs when $L_k = L_d$. This gives the value of the one-soliton solution amplitude of a single nonlinear Schrödinger equation

$$A_{\text{NLS}} = \sqrt{\frac{\sigma n\lambda_0}{\pi t_0^2 \chi_{\text{eff}}}} \approx 0.5 \times 10^4\ \text{esu}$$

for the 0.1 ps pulse duration. The corresponding length scale is $L_k^{(\mathrm{NLS})} \approx 70$ cm. The nonlinear Schrödinger one-soliton peak intensity can be estimated as

$$I_{NLS} = \frac{c(A_{NLS})^2}{8\pi} \approx 4 \times 10^9 \text{ W/cm}^2.$$

For comparison, the amplitude of a 0.1 ps 2π-pulse is $A_{2\pi} = 2\hbar d^{-1} t_0^{-1} \approx 4 \times 10^6$ esu. The peak intensity of the pulse reaches the magnitude $I_{2\pi} \approx 2 \times 10^{15}$ W/cm^2.

Another balance equality $L_k = L_c$ yields the electric field strength $A_c = (2n\Delta n \chi_{\mathrm{eff}}^{-1})^{1/2}$ known as the characteristic light wave field for a continuous wave (cw) of a nonlinear directional coupler [21], $A_c \approx 10^4$ esu, the intensity $I_c \approx 1.5 \times 10^{10}$ W/cm^2. This broadly means that for the input field amplitude values higher than A_c, the nonlinear birefringence initiated by Kerr processes begins to have a noticeable effect.

The quantity

$$L_r^{(2\pi)} = \frac{n\hbar\lambda_0}{2\pi^2 n_a d^2 t_0} \approx 5 \times 10^2 \text{ cm}$$

is the distance in the sample over which the reciprocal reaction of the medium in the form of polarization and population differences develops to produce coherent transients, e.g., the self-induced transparency [13], photon echoes [22, 23], optical nutations, breather waves [24], etc. For signals with a small pulse area θ [24],

$$\theta = d\hbar^{-1} \int_{-\infty}^{\infty} R(z, t)\, dt,$$

the parameter $L_r^{(2\pi)}$ serves as the absorption length. The pulse area of the NLS soliton $\theta_{NLS} = \pi d\hbar^{-1} t_0 A_{NLS} = 3 \times 10^{-3}\pi$ is extremely small in comparison with $\theta_{SIT} = 2\pi$.

4. EVOLUTION OF POLARIZATION STATES IN A FIBER. NUMERICAL ANALYSIS

We can now proceed to examine typical numerical results. We focus on the diagnostic of the temporal profile of the field amplitude and polarization parameters ε and θ at every cross-section of the nonlinear birefringent fiber. We assume the light wave to be in exact resonance with the homogeneously broadened atomic transition, i.e., $\nu = 0$. In order not to overcomplicate the problem, we also ignore the walk-off effect in this paper, although we observed some of its obvious results in our preliminary computations. In the numerical simulations demonstrated in Figs. 1–5 below, amplitudes (11) of the input pulses were chosen as $e_{m1} = \sqrt{3}/2$ and $e_{m2} = 1/2$ while the respective initial phases were 0 and π. We also set $f = 0.0015$, thus assuming that the resonance interaction process is not a strong perturbation to the fiber effects.

The propagation of a light pulse in a birefringent fiber is accompanied by the two-way coupling between the orthogonal counter-rotating polarization modes with the spatial beat period $2\pi\Delta\beta^{-1}$. No dispersion is involved in the numerical simulation at this stage. For the linear undoped fiber (i.e., when the contribution of the Kerr self- and cross-modulation effect can be neglected) the solution of Eqs. (7) and (8) is quite simple (Fig. 1a, 1b). The period of the partial energy transfer between the modes is $l_b = \pi l_c$. It is seen from the 3D plot of the azimuth θ and the ellipticity ε (Fig. 1c, 1d) that both functions are uniform across the pulse and oscillate in the course of propagation inside the fiber [18]. It is worth noting that if the launched pulse amplitudes were $e_{m1} = 1$, $e_{m2} = 0$, the azimuth angle θ would change from $-\pi/4$ to $\pi/4$ and the polarization state would change from the linear polarization ($\varepsilon = 0$) to a circular polarization of the opposite direction ($\varepsilon = \pm 1$).

When both polarizations are excited in an asymmetric manner, the ellipticity oscillates between the elliptical clockwise and elliptical anticlockwise polarizations. This is clearly seen from the gray scale modular surface of $\theta(\zeta, \tau)$ and $\varepsilon(\zeta, \tau)$ (Figs. 1e, 1f). The dark gray up to black corresponds to the maxima of the plotted function, while the light gray down to white, to the minima. The phase trajectories on the ε vs θ plane (with ε and θ calculated at the moments of peak intensity of the pulse) parameterized by ζ are closed circles (Fig. 1g). In this picture, each trajectory is associated with a different $\eta = (e_{m1} e_{m2}^{-1})^2$ ratio. The outer curve pertains to $\eta = 999$. The subsequent cycles correspond to $\eta = 99$, 9, 3, 1.5, 1.22. The biggest circle is the ultimate trajectory related to a nearly net circularly right-hand polarized light and small circles correspond to a nearly linearly polarized light. The circle in open dots corresponds to the case that was numerically investigated: $e_{m1} = \sqrt{3}/2$ and $e_{m2} = 1/2$. This numerical picture is in good agreement with the one presented in [18] for the cw-waves.

With the Kerr and walk-off effects ignored, the combined action of the linear birefringence ($l_c = 0.25$) and dispersion ($l_d = 1.0$) provides a well interpretable effect of the intensity hump spreading (Figs. 2a, 2b) in the depth of the fiber, as is clearly seen in the gray scale map (Fig. 2c, 2d). In this case, the polarization properties of the travelling field (Figs. 2e, 2f) are quite similar to those in Fig. 1. The spikes on both sides of the central area in Figs. 2e and 2f are the result of numerical fluctuations provoking random switchovers of the ellipticity ε and the azimuth θ on the wings of the propagating pulse, where the field is extremely weak in both polarizations. In Fig. 2g, we display the phase plane (ε vs. θ) for the coupling + dispersion case for the parameters $\eta = 999$, 9, 3, 1.22. The value $\eta = 3$ corresponds to the

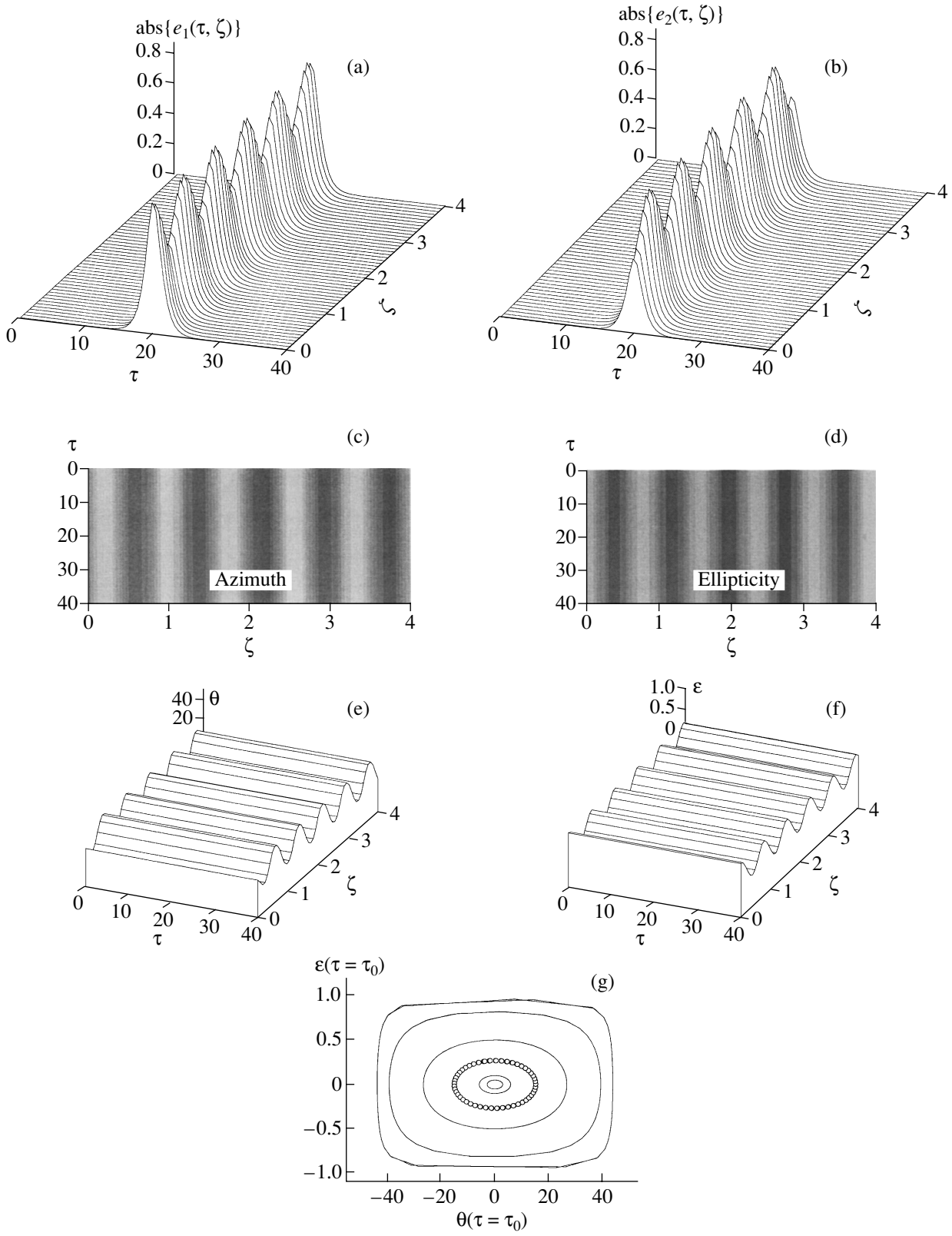

Fig. 1. (a, b) Normalized field strengths in counter-rotating polarization modes in birefringent ($l_c = 0.25$), linear ($l_k = \infty$), and dispersionless ($l_d = \infty$) fiber; (c, d) space–time evolution of the azimuthal angle θ and the ellipticity ε; (e, f) gray scale surface of the functions $\theta(\tau, \zeta)$ and $\varepsilon(\tau, \zeta)$; (g) phase trajectories of θ vs. ε taken at $\tau = \tau_0$ for different $\eta = (e_{m1}e_{m2}^{-1})^2$ ratios (see the text).

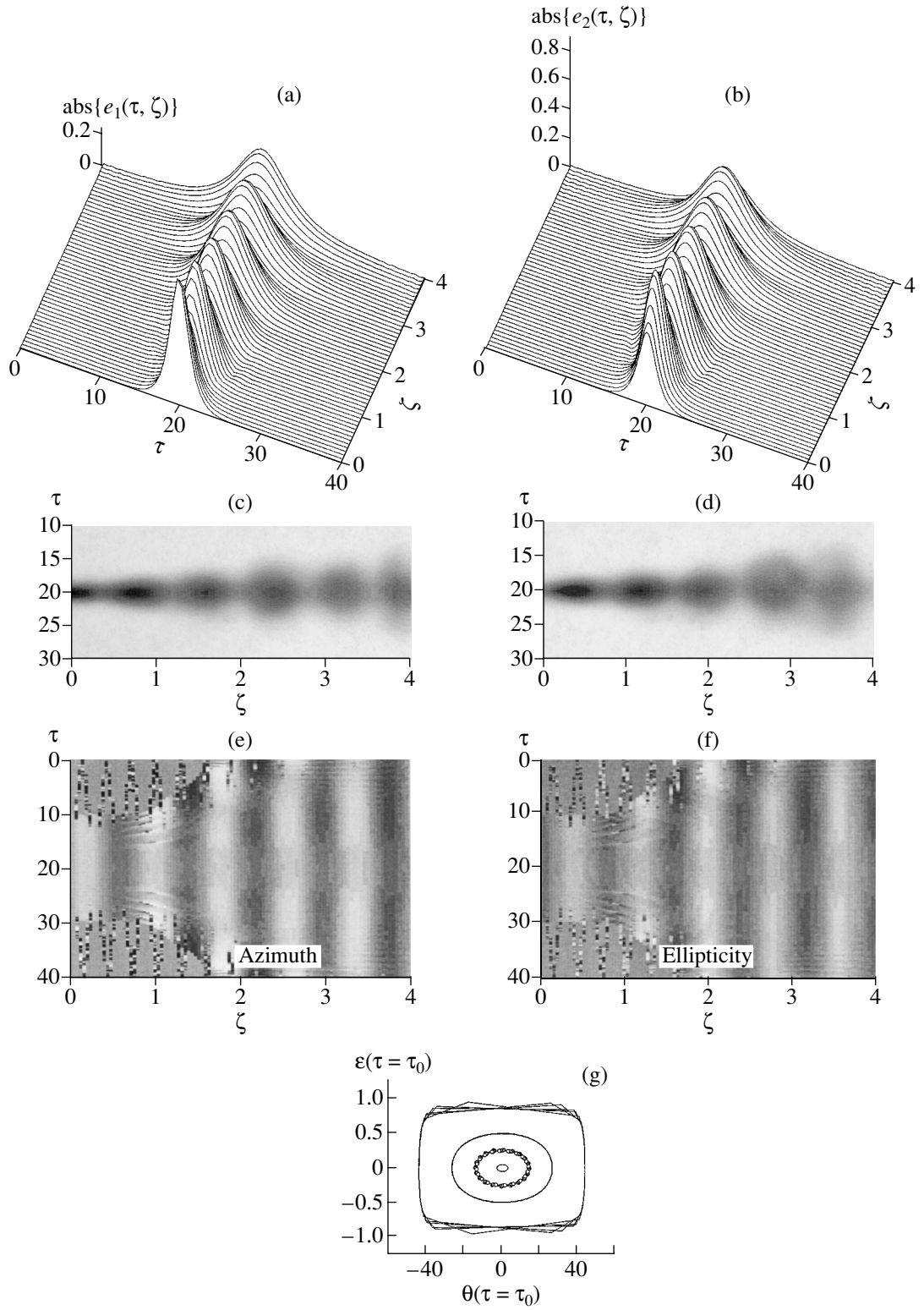

Fig. 2. (a, b) The same as in Fig. 1 for the parameters chosen as $l_k = \infty$, $l_c = 0.25$, and $l_d = 1.0$; (c, d) gray scale maps of the central parts of pictures (a, b), respectively; (e, f) gray scale maps of the azimuthal angle θ and the ellipticity ε; (g) the same as in Fig. 1.

case under numerical simulations. The plotted curves are quite similar to those in Fig. 1g.

The interplay between the linear coupling and the Kerr nonlinear phase modulation yields the picture that was not immediately evident (Fig. 3). We injected the pulses of the counter rotating polarization in the fiber, with the amplitudes of the pulses expressed in physical units satisfying the conditions $A_{m1} \approx 2A_c$ and $A_{m2} \approx A_c$.

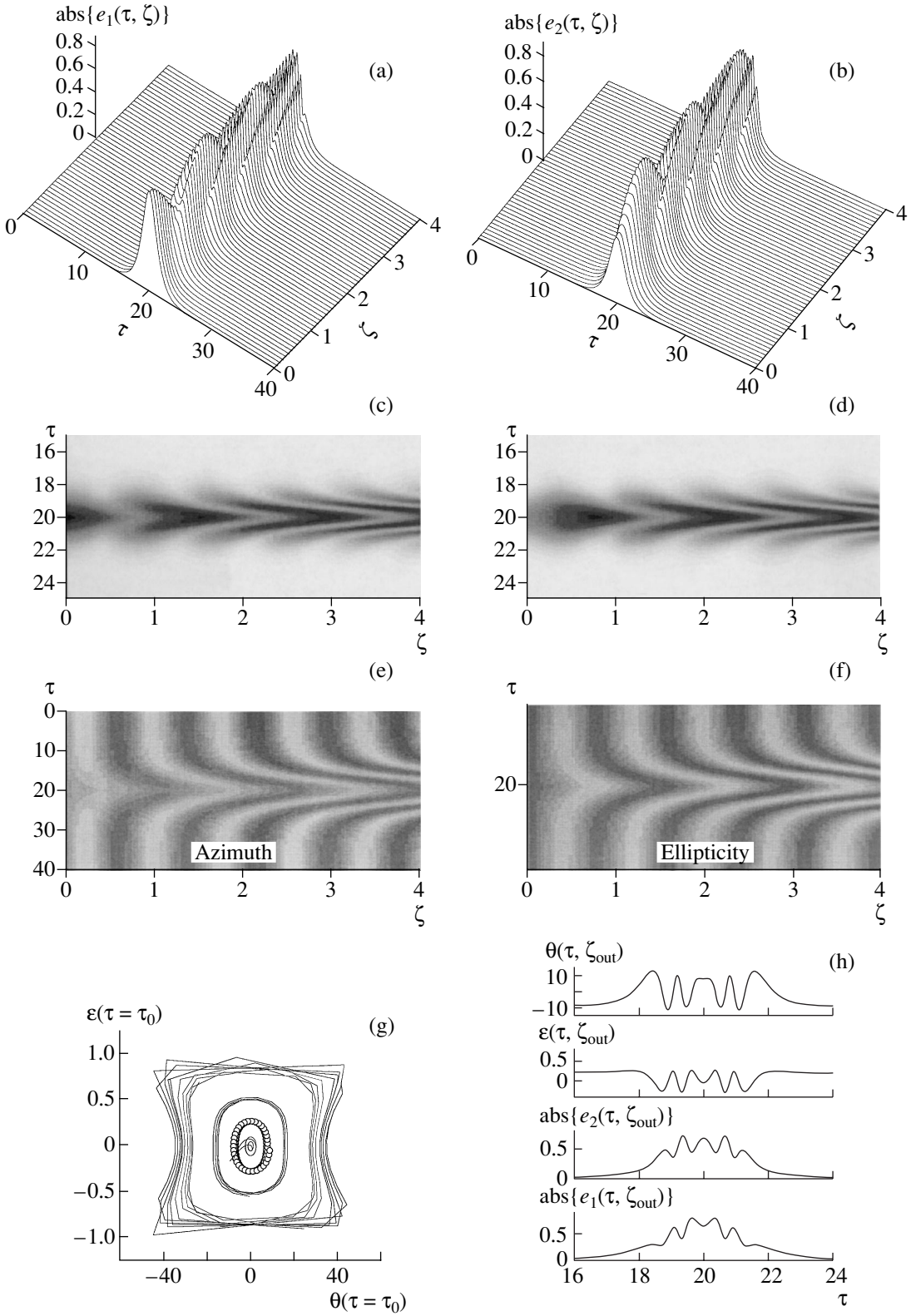

Fig. 3. (a, b) Normalized field strengths in polarization components of the pulse propagating in birefringent nonlinear and dispersionless fiber with $l_c = 0.25$, $l_k = 0.05$; (c–f) the same as in Fig. 2; (g) the same as in Fig. 1; (h) from top to bottom: the azimuthal angle θ, the ellipticity ε, and the polarization mode moduli at the exit of the fiber.

This corresponds to choosing $l_c = 0.25$ and $l_k = 0.05$ for the characteristic lengths. In this case, one could expect the Kerr to compress the energy couples back to the mode from the conjuncted polarization state every time. Instead, we observe the interference between the coupling processes with different beat periods. The coupling is revealed in the form of the inserted cycles when every new growth of the amplitude begins while the previous one has not yet finished. The physical explanation may be found if one notes that both input amplitudes are chosen to be of the order of the critical strength of the electric field A_c for cw switching. For such intensities, the Kerr processes become sufficiently strong to make the birefringence a nonlinear process and the beat period can even grow unlimitedly [25–28]. Attention must be drawn to the fact that the periodicity of the onsets of the back and forth coupling cycles approaches the value prescribed by the choice $l_c = 0.25$ (compare with Fig. 1). The linear behavior occurs only on the slopes of the pulse envelope, where the field intensity has not reached the critical value. The further growth of the pulse field strength in a pulse envelope forces the beat period to increase as well. The result is seen in Figs. 3a, 3b and 3c, 3d showing the 3D picture and the gray scale map of the polarization mode dynamics, respectively. The envelopes of the field in both polarization modes experience a temporal counter-phase modulation in the central part of the propagating waveform (Fig. 3h). The modulation of the amplitudes of the counter rotating polarization modes leads to the oscillation of ε and θ over τ in the propagating light wave that is clearly seen from the gray scale maps in Figs. 3e and 3f and on the comparative plots of the fields, ellipticity, and azimuth at the exit from fiber placed in Fig. 3h. Figure 3g shows the (ε, θ) phase plane for the same values of the parameter η as above. Unlike the previous cases, the ultimate circle is distorted, which agrees with the analysis in [18]. This cycle is smeared because the spatial modulation of the peak intensity of polarization components is complex.

Our calculations presented in Fig. 4 illustrate the combined action of the linear birefringence, Kerr nonlinearity, and dispersion. The dispersion length $l_d = 1$ serves as a scale length, while the coupling length l_c and the Kerr length l_k are shorter, $l_k = 0.1$ and $l_c = 0.25$. The choice of parameters dictates the values of the polarization mode amplitudes at the entrance to the fiber: $A_{m1} \approx \sqrt{2}\,A_c \approx 2.7A_{NLS}$ and $A_{m2} \approx 0.8A_c \approx 1.6A_{NLS}$. Weak ripples at the edges of the computational grid are due to the time boundary conditions.

The current case is not a completely integrable problem because of the intermode coupling. The propagating pulse cannot find a stable form at least over the distance considered here. In one of our preliminary computations under the conditions similar to those in Fig. 4, but with $l_k \approx 0.05$ (i.e., for a greater amplitude), we observed the breaking up of the input pulses of both polarizations into two separate subpulses subsequently scattering aside.

The periodical squeezing of the pulse shape, a feature of a high-order NLS solution, produces new oscillations on the wings of the pulse (Figs. 4a, 4b) because the Kerr processes and dispersion are spatially mismatched. The dispersion spreading is noticeable at several ($\zeta \sim 4$) normalized lengths when the dispersion chirp fills almost the entire time window (Figs. 4c, 4d). It is then natural that polarization properties of the light wave (i.e., the alternation of dark and light shades of the gray) map the broadening area of the spatial-temporal oscillation of the polarization components (Figs. 4e, 4f) caused by the dispersion, thereby making the entire picture rather complicated. As in Fig. 3h, the oscillations of the field remain out of phase in polarization modes (Fig. 4g). We note that there are fewer coupling periods in Figs. 4a and 4b than in Fig. 1 or Fig. 2. Clearly, the nonlinear narrowing and peak amplification drive the propagation of the pulse into a nonlinear birefringent regime. A further growth of the pulse input amplitudes strengthens the inequality $l_k < l_d$, thereby making the process somewhat analogous to that in Fig. 3, plus the dispersion-originated oscillations spreading away of the sharp central peak.

The resonance interaction of a short pulse with the ensemble of resonance atoms is now added to the conventional fiber effects as indicated in Eqs. (7), (8) and (12)–(14). The evolutionary behavior of the counter circularly polarized components with the input amplitudes $A_{m1} \approx 2A_c \approx 4A_{NLS}$, $A_{m2} \approx A_c \approx 2A_{NLS}$ is plotted in Figs. 5a, 5b and 5c, 5d. We assume the resonance interaction to be weak by setting $f = 0.0015$. Under this condition, the population differences insignificantly deviate from their initial values. The spatial scale of the process is $l_d = 1.0$, whereas $l_k = 0.05$, $l_c = 0.25$, and $l_r = 0.01$. The value of the resonance interaction length $L_r^{(2\pi)}$ can be estimated as $L_r^{(2\pi)} \approx 7L_d$ (see (6)). This means that the total length of the fiber in Fig. 5 is about $0.6\,L_r^{(2\pi)}$ or $4L_d$. The resonance interaction process transfers energy more effectively than the dispersion off the pulse to the radiation born by the reciprocal reaction of the medium in the pulse-after-action region. It is then clear that in comparison with Fig. 4, the amplitudes of the humps rapidly decrease in the propagation direction (Figs. 5a, 5b).

Attention should be drawn to two humps in the center of Fig. 5g. These are the above-mentioned relics of the NLS *N*-soliton break up. The visible asymmetry of the pattern relative to the initial pulse position results from the delayed response of the resonance subset. Generally, we can predict that at longer distances inside the doped fiber, the well-evolved dispersion and coherent "ring" effects can hardly be distinguished.

The polarization properties of the light pulse are displayed in the gray scale maps in Figs. 5e and 5f. It is interesting to note that these pictures preserve the peri-

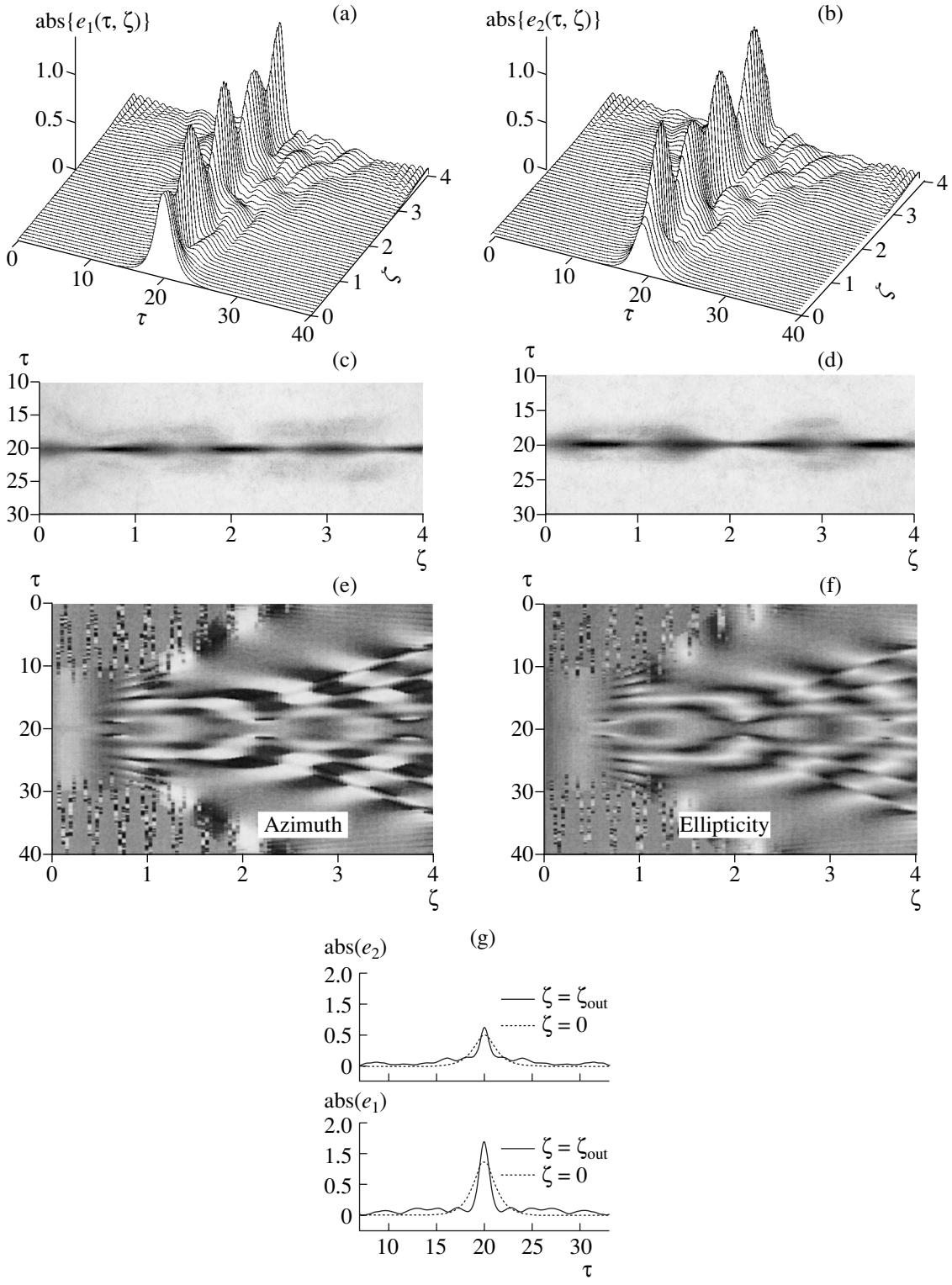

Fig. 4. (a, b) Normalized field strengths in polarization components of the pulse for $l_c = 0.25$, $l_d = 1.0$, and $l_k = 0.1$; (c–f) the same as in Fig. 2; (g) polarization mode shapes (absolute values) at the entrance (dashed line) and at the exit (solid line) of the fiber.

odic alternation of the regions with the opposite ellipticity and azimuthal angle owing to the linear coupling (see Figs. 2e, 2f). In our further computations (not shown), when we set $l_r = 0.001$ for the ten times larger concentration of impurities, we saw the resonance oscillations already filling the entire (ζ, τ) computational area at an early stage of the pulse propagation. It was interesting to observe how the increase of the

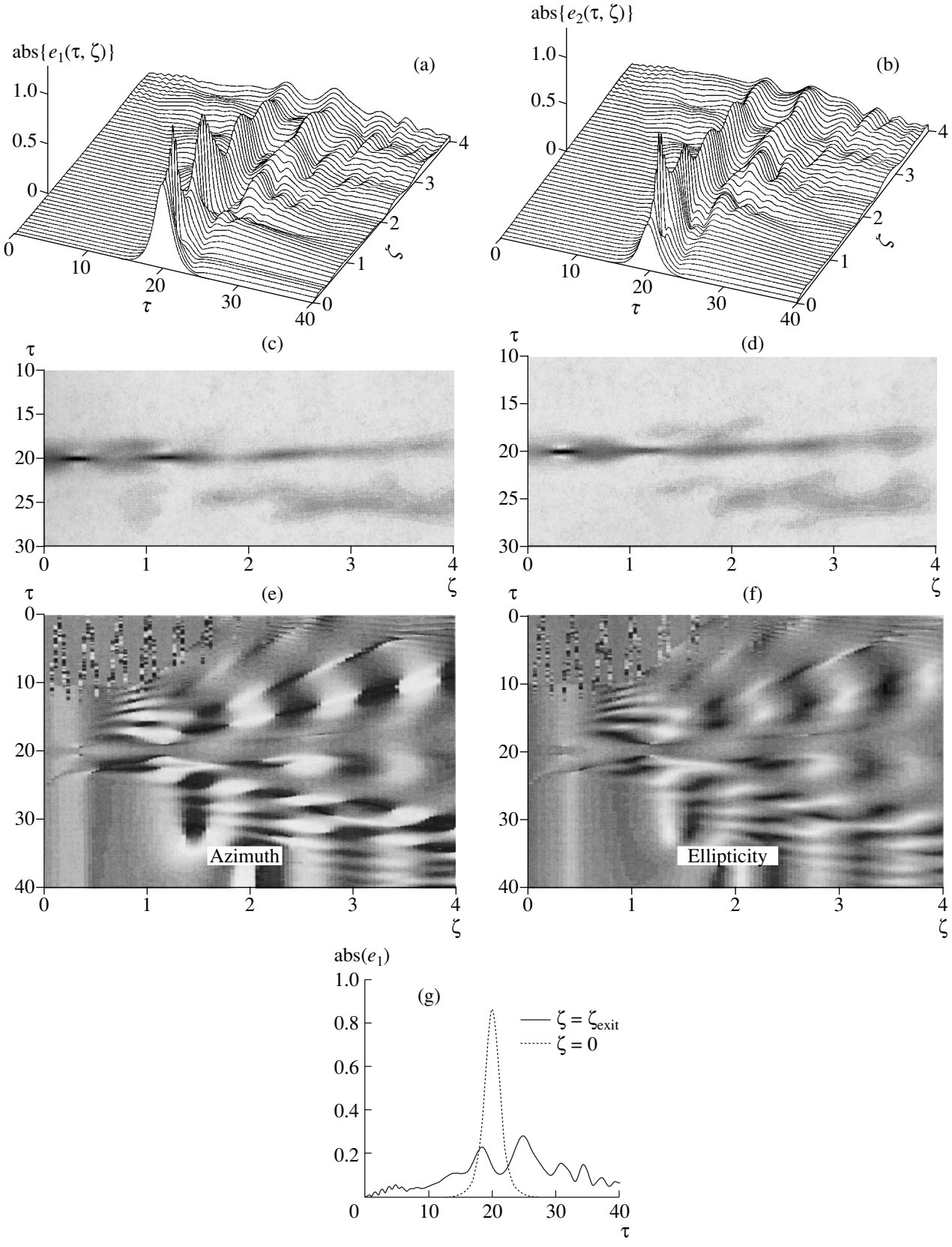

Fig. 5. (a, b) Normalized field strengths in polarization components of the pulse propagating in a fiber with impurities ($l_r = 0.01$). Other parameters are $l_c = 0.25$, $l_d = 1.0$, and $l_k = 0.05$; (c–f) the same as in Figs. 2–4; (g) polarization modules at the entrance (dashed line) and at the exit (solid line) of the fiber.

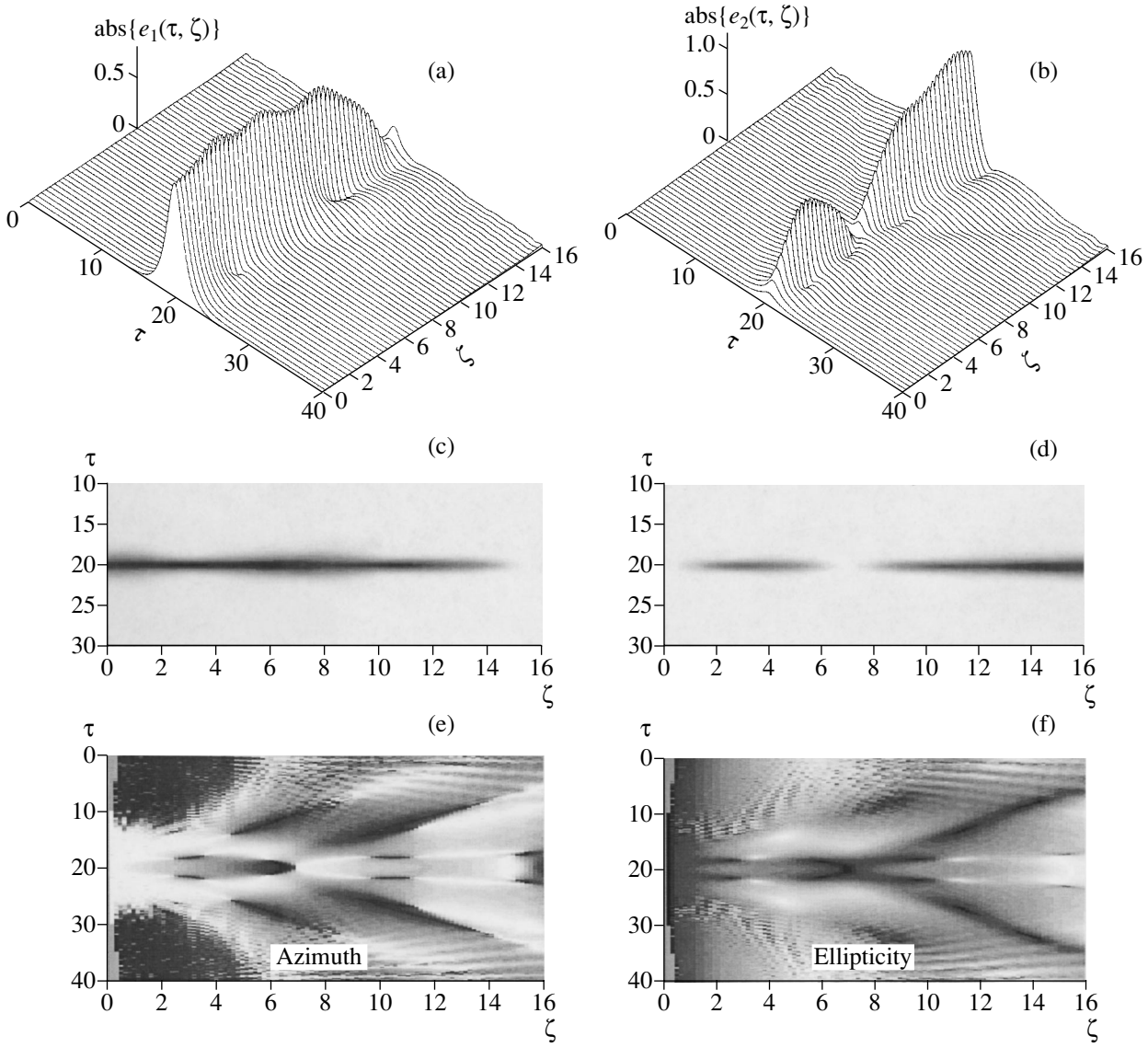

Fig. 6. (a–f) The same as Figs. 4a–4f with the parameters $l_d = 2.0$, $l_c = 4.0$, $l_k = 0.33$.

dopant concentration developed the generic picture of the periodic azimuth and ellipticity variations with the same beat period seen in Figs. 1 and 2. Qualitatively, this can be regarded as a result of the resonance absorption when the progressive damping of the field humps decreases the field amplitude below the critical value of the electric field strength A_c, thereby driving the process back into the linear regime, when l_c begins to be shorter than l_k. Anyway, because the dispersion and the resonance interaction are time-dependent processes, they introduce temporal modulation to the basic polarization picture consistent with the linear birefringence.

The comparison of our numerical simulation with the known results [25] is displayed in Figs. 6 and 7. The parameters introduced in this paper correspond to the analogous quantities in [25, Fig. 1b] if we set $l_d = 2$, $l_c = 4$, $l_k = 0.33$, and $e_{m1}^2 = 1.25$. This yields the estimate $A_{m1} \approx 4A_c \approx 3A_{NLS}$, $A_{m2} = 0$ in physical units for the input field amplitudes in the cases depicted in Figs. 6 and 7. In both pictures, we kept the original [25] length of the fiber, although it corresponds to $l_{\text{fiber}} = 8l_d$ in our conventions. The discrepancy originates from renormalizing the factor in the dispersion term in (5) by the coefficient 1/2. In Fig. 6, we reproduce the results of [25] observing the formation of two distinct periodicities of the coupling process between the modes. The pulse shape dynamics (Figs. 6a, 6c and 6b, 6d) can be physically interpreted in terms of the l_c/l_k ratio, which in the current case is the biggest of all those described above, $l_c/l_k = 12$. The coupling process between the modes must then reveal a nonlinear behavior because of the power dependence of the energy exchange period. In

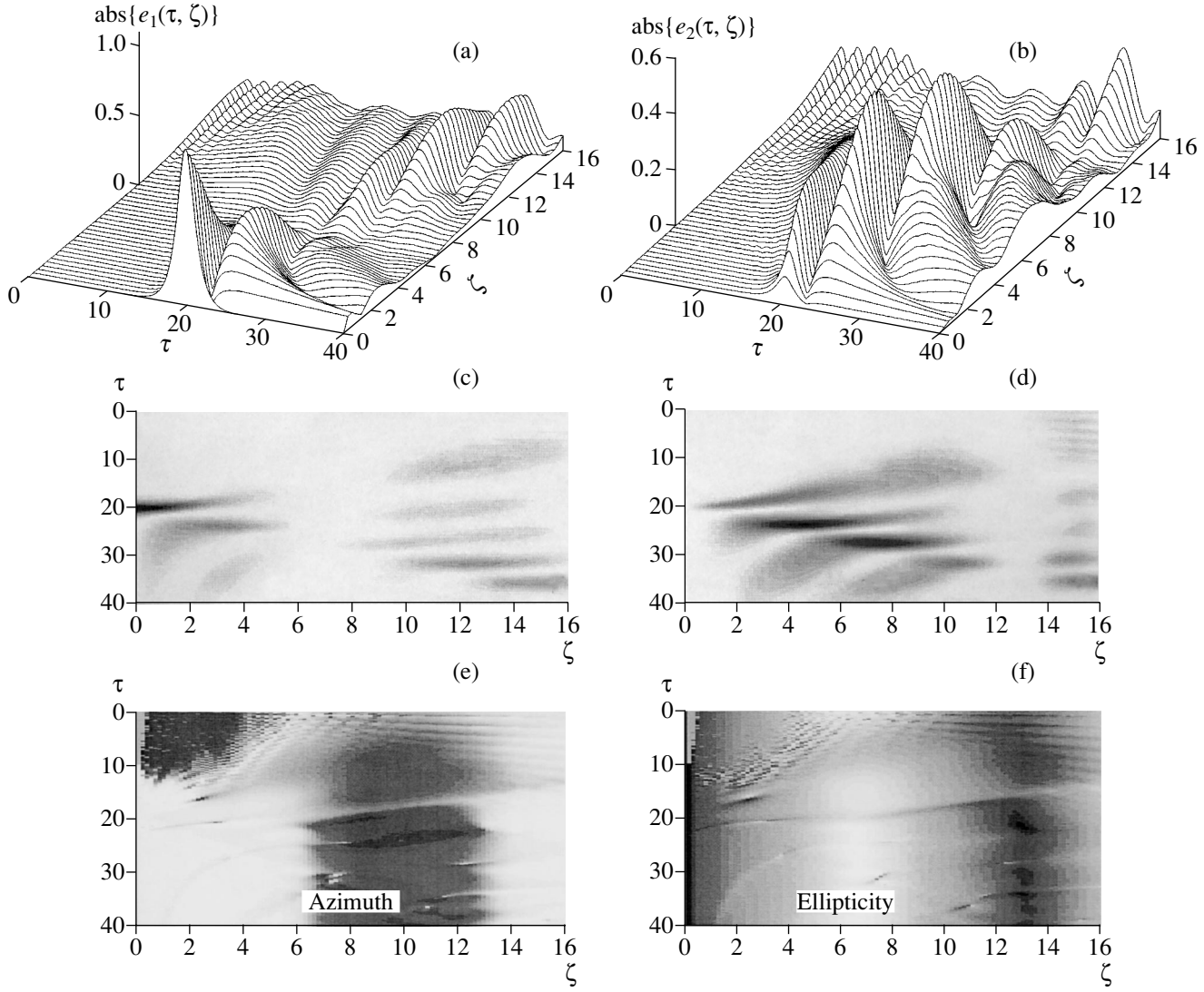

Fig. 7. (a–f) The same as Figs. 5a–5f with the parameters $l_d = 2.0$, $l_c = 4.0$, $l_k = 0.33$, $l_r = 0.01$.

the L_d units, the linear coupling length (i.e., at low power) should have been $l_b = L_b L_d^{-1} = \pi L_c L_d^{-1} = 2\pi$. As a matter of fact, the visual estimate of the beat period in Figs. 6a, 6b and 6c, 6d yields $l_b^{nl} = L_b^{nl} L_d^{-1} \approx 4.5\pi > l_b$. We note that in the case under consideration, the period of the typical higher-order soliton compressions became power dependent. This manifests the difference between our approach and the ideal completely integrable model [29].

Figures 6e and 6f shows the gray scale (τ, ζ) maps of the azimuthal angle θ and the ellipticity parameter ε. It is seen that the time–pace features of the dispersion process are distinctly reproduced. The brightest lines and spots in the region of pulse slopes demonstrate abrupt changes of the polarization state due to a rapid growth or drop of the field in one polarization mode compared to the other.

The reciprocal reaction of resonance impurities on the field of the propagating short pulse can noticeably change the space-time picture of polarization dynamics in the analysis of nonlinear effects in a pure fiber by Trillo *et al.* [25]. For the computational variant presented in Fig. 7, all the fiber parameters and input pulse amplitudes remain unaltered with respect to the case in Fig. 6. But in contrast to the variant in Fig. 6, the field of a short propagating pulse is now coupled to the resonance subsystem, and therefore, the complete system in Eqs. (7)–(8) and (12)–(14) must be solved numerically. The value of the normalized resonance length was set as $l_r = 0.01$, with the corresponding physical length of the resonance interaction $L_r^{(2\pi)} \approx 3L_d$ and the parameter $f = 0.0015$. The population difference between the resonance levels (which was computed but is not shown here) remains practically unchanged because the areas of the coherent pulses are small.

The resonance response changes the waveform of the polarized light. In the current case, the intensity of the interaction process is higher than in Fig. 5 because the resonance interaction length is shorter. The energy of the input pulse is rapidly transferred to oscillations of the resonance polarization (Figs. 7a, 7b) in the region of the retarded action of the propagating pulse (Figs. 7c, 7d). The resonance absorption noticeably weakens the humps for greater ζ, moving them out of the calculation window. It is interesting that the intensity damping leads to the restoration of the linear beat period $l_b \approx 2\pi$, and the propagation process is therefore converted from a nonlinear regime to the linear one. Thus, the nonlinear phase modulation and dispersion do not play the leading role in the dynamics of the pulse. This explains why the maps of polarization parameters (Figs. 7e, 7f) are so flat. In fact, the slow variations in the form of dark and light stripes due to the linear coupling process period (Figs. 7e, 7f) are the only prominent feature of the displayed plots. It is interesting that the oscillations of the field envelope caused by the retarded reciprocal reaction of the medium, being in phase, do not produce a modulation of the polarization parameters ε and θ except in the vicinities of especially rapid changes of the field.

Passing to a higher concentration ($l_r = 0.001$) demonstrates the typical features of the coherent phenomena in a resonance medium. In the depth of the fiber, there is no solitary wave; instead, we have a waveform with the oscillating envelope. This wave packet rapidly shifts towards the later times, leaving the calculation grid somewhere at $\zeta = 6$. The polarization properties remain indifferent to the complete destruction of the pulse and the linear birefringence beat period is preserved.

5. CONCLUSION

In this paper, we have tried to give an indication of a rich space time dynamics arising from the propagation of an elliptically polarized light pulse in a nonlinear birefringent doped fiber. The resonance impurities in the form of two-level atoms were included in the model in addition to the full set of nonlinear fiber effects. We have concentrated on the case of a weak input field, for which the amplitude of the pulse is about the amplitude of a single NLS pulse and the coupling to the resonance system is therefore not strong. A trivial account of the weak effect of the resonance system on the propagating optical pulse leads to a linear absorption. Generally speaking, the coherent interaction of short pulses with resonance atoms is a non-Markovian process [24, 30]. Moreover, the degeneration of resonance levels gives the contribution to birefringence that is nonlocal in time. With the exception of big detunings off the resonance, the analytical consideration of all these effects is extremely difficult. Therefore, a direct numerical simulation of the pulse evolution is preferable. But even within the weak-interaction approximation, the general picture proved to be sufficiently complex. The polarization properties of the pulsed light are nonstationary across the pulse width and can also drastically change in space. Our numerical simulations show that the polarization dynamic is basically featured by the interplay between the Kerr nonlinear self- and cross-phase modulation and dispersion, while the linear birefringence leads to a spatial modulation of the azimuthal angle and the ellipticity. There is a range of the input amplitudes where the birefringence becomes a nonlinear power dependent process because of the Kerr cross-phase modulation, and the power beat period can therefore grow. At the same time, when the spatial scale of the resonance interaction becomes less than or about the characteristic lengths of the fiber effects, the propagating pulse experiences a strong distortion and a resonance absorption. The intensity damping leads to the restoration of the linear beat period, there by converting the propagation process from the nonlinear regime to the linear one. To observe perceptible coherent effects such as the SIT or the photon echo, one must take many times more powerful pulses, which in their turn excite a higher-order N-soliton effect. Therefore, the problem requires a special treatment.

ACKNOWLEDGMENTS

We are grateful to S. Kozlov, V. Sazonov, and V. Kozlov for fruitful discussions.

REFERENCES

1. G. P. Agrawal, *Nonlinear Fiber Optics* (Academic, San Diego, 1989).
2. A. I. Maimistov and A. M. Basharov, *Nonlinear Optical Waves* (Kluwer, Dordrecht, 1999).
3. A. Hasegawe, *Optical Solitons in Fibers* (Springer-Verlag, Berlin, 1990).
4. K. Nakagawa, Sh. Nishi, K. Aida, and E. Yoneda, J. Lightwave Technol. **9**, 198 (1991).
5. A. I. Maimistov and A. M. Basharov, Izv. Akad. Nauk, Ser. Fiz. **62**, 354 (1998).
6. K. Porsezian and K. Nakkeeran, Phys. Rev. Lett. **74**, 2941 (1995).
7. S. L. McCall and E. L. Hahn, Phys. Rev. **183**, 457 (1969).
8. M. Nakazawa, E. Yamada, and H. Kubota, Phys. Rev. Lett. **66**, 2625 (1991).
9. A. D. Boardman and G. S. Cooper, J. Opt. Soc. Am. B **5**, 403 (1988).
10. A. I. Maimistov and E. A. Manykin, Zh. Éksp. Teor. Fiz. **85**, 1177 (1983) [Sov. Phys. JETP **58**, 685 (1983)].
11. M. Horowetz and Y. Silberberg, Opt. Lett. **22**, 1760 (1997).
12. H. Steudel, J. Mod. Opt. **35**, 693 (1988).
13. A. I. Maimistov, A. M. Basharov, S. O. Elyutin, and Yu. M. Sklyarov, Phys. Rep. **191**, 1 (1990).
14. A. D. Boardman and G. S. Cooper, J. Mod. Opt. **35**, 407 (1988).

15. C. R. Menyuk, J. Opt. Soc. Am. B **5**, 392 (1988).
16. C. R. Menyuk, IEEE J. Quantum Electron. **QE-25**, 2674 (1989).
17. R. T. Taha and M. J. J. Ablowitz, Comput. Phys. **55**, 203 (1984).
18. H. G. Winful, Appl. Phys. Lett. **47**, 213 (1985); Opt. Lett. **11**, 33 (1986).
19. M. Nakazawa, Y. Kimura, K. Kurokawa, and K. Suzuki, Phys. Rev. A **45**, R23 (1992).
20. C. R. Menyuk, IEEE J. Quantum Electron. **QE-23**, 174 (1997).
21. B. Diano, G. Gregori, and S. Wabnitz, Opt. Lett. **11**, 42 (1986).
22. V. L. daSilva and Y. Silberberg, Phys. Rev. Lett. **70**, 1097 (1993).
23. J. Hegarty, M. M. Broer, B. Golding, *et al.*, Phys. Rev. Lett. **51**, 2033 (1983).
24. L. Allen and J. H. Eberly, *Optical Resonance and Two-Level Atoms* (Wiley, New York, 1975).
25. S. Trillo, S. Wabnitz, E. M. Wright, and G. I. Stegeman, Opt. Commun. **70**, 166 (1989).
26. S. O. Elyutin and A. I. Maimistov, Chaos, Solitons and Fractals **11**, 1253 (2000).
27. D. J. Muraki and W. L. Kath, Physica D (Amsterdam) **48**, 53 (1991).
28. D. J. Kaup, B. A. Malomed, and R. S. Tasgal, Phys. Rev. E **48**, 3049 (1993).
29. S. V. Manakov, Zh. Éksp. Teor. Fiz. **65**, 505 (1973) [Sov. Phys. JETP **38**, 248 (1973)].
30. M. D. Crisp, Phys. Rev. A **1**, 1604 (1970).

Journal of Experimental and Theoretical Physics, Vol. 93, No. 4, 2001, pp. 753–759.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 863–870.

FLUIDS

New Phase Nucleation in Electric Fields

V. S. Vorob'ev* and S. P. Malyshenko

Institute of High Temperatures Scientific Association (IVTAN), Russian Academy of Science, Izhorskaya ul. 13/19, Moscow, 127412 Russia

**e-mail: vrbv@mail.ru*

Received December 22, 2000

Abstract—The formation of vapor bubbles in a superheated liquid or of liquid drops in a superheated vapor in a uniform electric field and at a charged center is considered. The work done for the formation of drops as well as their critical size decrease in all cases. The critical size of bubbles always increases, while the work decreases at a charged center and in a uniform field. An explanation is offered for the results of experiments on the initiation of boiling of a superheated liquid by electric field pulses. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The effect of external magnetic fields causing the polarization of a medium on the phase equilibrium conditions was considered by many authors [1–19]. It was found that the electric field facilitates the condensation of superheated vapor by reducing the critical size of liquid drops and the work required for their formation. However, the conclusions concerning vapor bubbles formed in a superheated liquid are contradictory. According to [11, 12], an electric field must hamper the nucleation of the vapor phase in the liquid. However, a qualitatively opposite effect was observed in some experiments: a pulsed electric field applied to a dielectric initiated the boiling of a superheated liquid [13–18]. In order to remove this contradiction, a number of hypotheses reflecting different opinions concerning the experimentally observed effect were put forth in [13, 16, 17]. These effects included the liberation of a gas at electrodes followed by Joule heating and liquid breakdown [13, 16] and the existence of an electric double layer at the liquid–cell-wall interface, which facilitates the formation of the gaseous phase [15]. In specially enacted experiments [14], electrodes were isolated from the liquid by a glass casing which suppressed the gas emission. Nevertheless, these experiments confirmed the intensification of boiling in an electric field. In the hypothesis of a double layer, the reasons behind its formation are not quite clear and the value of its charge remains undetermined. According to the authors of [19], the motion induced by electric forces may affect vapor formation.

In the present work, the nucleation of a new phase in a dielectric medium polarized by an external electric field is studied on the basis of the conservation laws at the interface and from an analysis of the expression for the excess work required for the formation of a nucleus. The nucleation at a charged center and in a uniform field is considered.

The surface forces emerging at the interface (i.e., the electric force directed towards a less dense medium and the Laplace force associated with the curvature of the surface and elevating pressure in a nucleus) are of considerable importance. For a liquid drop in the gaseous phase, these forces have opposite directions, which "unloads" a drop, effectively reducing its surface tension, and leads to a decrease in the size of a critical liquid phase nucleus in the supersaturated vapor. On the contrary, for a gas bubble in the liquid medium, the action of both forces is added so that both forces compress the bubble. As a result, the size of a critical nucleus in the presence of a field is larger than in zero field. However, the work done for the formation of a critical-size bubble decreases in a uniform field and in the nonuniform field of a charge center on account of the change in the electric field energy. The latter circumstance makes it possible to interpret the results of experiments in which the boiling of a superheated liquid is initiated by a nonuniform electric field.

2. CONDITIONS AT THE PHASE INTERFACE IN AN ELECTRIC FIELD

Let us consider the conditions which must be satisfied by the parameters of a dielectric polarized by an external electric field at a curvilinear stationary interface in phase equilibrium ($T_1 = T_2$, the mass flow through the interface is zero). The balance of forces at the interface leads to a relation for the pressure difference in the phases [4] and the energy balance leads to the equality of the total chemical potentials of the phases. This gives

$$p_1 - \frac{E_1^2}{8\pi}\left(\rho_1\frac{\partial\varepsilon_1}{\partial\rho_1}\right)_T + \Delta F - F_\sigma = p_2 - \frac{E_2^2}{8\pi}\left(\rho_2\frac{\partial\varepsilon_2}{\partial\rho_2}\right)_T, \quad (1)$$

$$\mu_1(p_1) - \frac{E_1^2}{8\pi}\left(\frac{\partial\varepsilon_1}{\partial\rho_1}\right)_T = \mu_2(p_2) - \frac{E_2^2}{8\pi}\left(\frac{\partial\varepsilon_2}{\partial\rho_2}\right)_T, \quad (2)$$

1063-7761/01/9304-0753$21.00

where p_1, p_2 and ρ_1, ρ_2 are the pressures and densities of the substances of the phases at the interface, E_1 and E_2 are the electric field strength at the interface, which are connected through the following boundary conditions:

$$E_1^2 = E_t^2 + E_n^2, \quad E_2^2 = E_t^2 + E_n^2\varepsilon_1^2/\varepsilon_2^2,$$

E_n and E_t being the normal and tangential components of the field strength in the first medium,

$$F_\sigma = \sigma\left(\frac{1}{R_1} + \frac{1}{R_2}\right),$$

σ is the surface tension, R_1 and R_2 are the principal radii of curvature of the phase boundary at the point under investigation. ΔF is the surface electric force associated with the difference in the permittivities of the media, is directed towards the medium with a smaller value of ε, and defined as [4]

$$\Delta F = \frac{\varepsilon_1 - \varepsilon_2}{8\pi}\left(E_t^2 + E_n^2\frac{\varepsilon_1}{\varepsilon_2}\right). \quad (3)$$

Formula (2) is convenient since in zero field it is transformed, together with formula (1), into the standard conditions of phase equilibrium known from the thermodynamics for surfaces with a curvature:

$$\mu_1(p_1) = \mu_2(p_2), \quad p_1 = p_2 + F_\sigma.$$

Linearizing (to within linear terms) the chemical potentials in Eq. (2) in pressure in the vicinity of point p_s corresponding to phase equilibrium in zero field and for zero curvature, we arrive at, instead of (1) and (2), the following relations containing no chemical potentials:

$$p_{1E} - p_{2E} = F_\sigma - \Delta F, \quad (4)$$

$$(p_s - p_{2E})\left(1 - \frac{v_2}{v_1}\right) = F_\sigma - \Delta F, \quad (5)$$

where

$$p_{2E} = p_2 - \frac{E_2^2}{8\pi}\left(\rho_2\frac{\partial\varepsilon_2}{\partial\rho_2}\right)_T,$$
$$p_{1E} = p_1 - \frac{E_1^2}{8\pi}\left(\rho_1\frac{\partial\varepsilon_1}{\partial\rho_1}\right)_T, \quad (6)$$

and $v = 1/\rho$ is the specific volume of the substance.

In the presence of an electric field, the quantity $p_{2E} = p_0$ approximately preserves its value which is equal to pressure p_0 a point where the field is equal to zero [4]. Consequently, the contribution of the electrostriction forces to the resultant pressure drop at the interface is approximately equal to zero. This circumstance allows us to disregard these forces at the very outset as it was done, for example, in [5, 6].

3. NUCLEATION

Let us suppose that one of the phases has a small size and is formed within the main extended phase which is in a metastable state in the presence of an electric field. We are speaking of the formation of liquid drops in a supercooled vapor or of vapor bubbles in a superheated liquid. We will consider this problem for a nonuniform centrosymmetric field and for a uniform field.

3.1. Nucleation at a Charged Center

If the charge Q creating the field is located at the center of a bubble (for a liquid drop, this problem is considered in [5]), the problem becomes spherically symmetric and the nucleus is a sphere of radius R. We are not interested in the density distribution over the volume of the nucleus and, hence, assume that it is constant.

Since the temperature and the total chemical potential of the nucleus coincide with the values of these quantities for the ambient, the work required for its creation is determined by the change in the thermodynamic potential $\Omega = -pV$ of the substance and by the change in the electric field energy. We will write these quantities in the form

$$\Delta\Omega = 4\pi R^2\sigma - \frac{4\pi R^3}{3}\left[\left(p_1 - p_2 + \frac{Q^2}{8\pi R^4}\right)\left(\frac{1}{\varepsilon_1} - \frac{1}{\varepsilon_2}\right)\right]. \quad (7)$$

It is noteworthy that the contribution from the change in the electric field energy for a bubble or drop appears in Eq. (7) with different signs.

We can find the extrema by differentiating Eq. (7) with respect to R taking into account the dependence of the permittivity of the medium on density. This gives

$$\frac{2\sigma}{R} = p_1 - \frac{Q^2}{8\pi\varepsilon_1^2}\left(\rho_1\frac{\partial\varepsilon_1}{\partial\rho_1}\right)_T - p_2 + \frac{Q^2}{8\pi\varepsilon_2^2}\left(\rho_2\frac{\partial\varepsilon_2}{\partial\rho_2}\right)_T - \frac{Q^2}{8\pi R^4}\left(\frac{1}{\varepsilon_1} - \frac{1}{\varepsilon_2}\right). \quad (8)$$

It can be seen that we arrive precisely at condition (1) derived from the conservation laws and written for the spherical case. In zero field, expression (8) gives the Laplace formula for the radius of a critical nucleus. The equality of the total chemical potentials at the interface allows us to go over from Eq. (8) to a formula similar to Eqs. (5):

$$(p_s - p_0)\left(1 - \frac{v_2}{v_1}\right) = \frac{2\sigma}{R} + \frac{Q^2}{8\pi R^4}\left(\frac{1}{\varepsilon_1} - \frac{1}{\varepsilon_2}\right). \quad (9)$$

In order to analyze expressions (7) and (9), it is convenient to use the dimensionless units $p' = p/p_0$ and $r = R/R_c$, where R_c is the size of the critical nucleus in zero

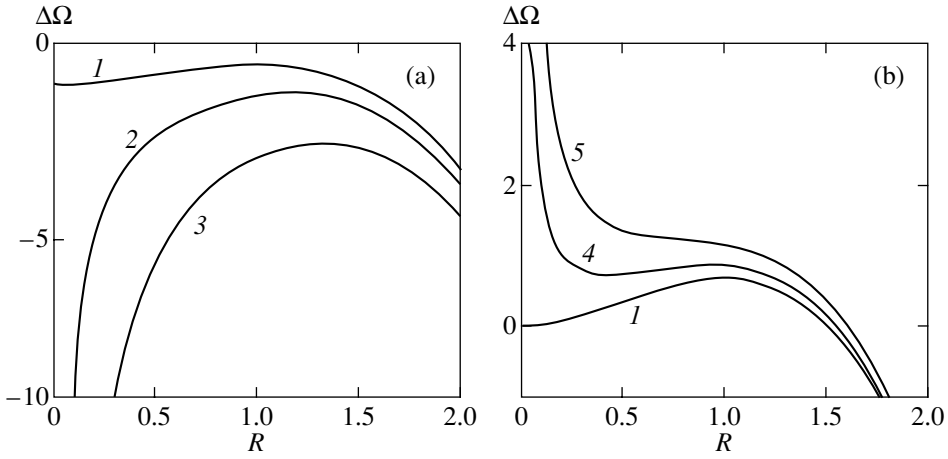

Fig. 1. Dependence of the work of nucleation on the field of a charged center on the radius: (a) bubble; (b) drop. Curves correspond to zero field (*1*), $F =$ (*2*) 1, (*3*) 3, (*4*) 0.2, and (*5*) 0.5.

field. In these units, Eq. (9) assumes the form (primes are omitted)

$$S\left(1-\frac{1}{r}\right) = \pm\frac{F}{r^4}, \tag{10}$$

where

$$F = \frac{Q^2}{8\pi R_c^4 p_0}\left|\frac{1}{\varepsilon_1}-\frac{1}{\varepsilon_2}\right|$$

characterizes the field and the quantity

$$S = \frac{p_s - p_0}{p_0}\left(1-\frac{v_2}{v_1}\right)$$

is the effective degree of supersaturation. For a vapor bubble in a superheated liquid, $p_s/p_0 \gg 1$ and $v_2/v_1 \ll 1$; consequently, parameter S is large. For a drop in a supersaturated vapor, it is also large, but now due to the fact that $p_s/p_0 < 1$ and $v_2/v_1 \gg 1$. In Eq. (10), the upper (plus) sign should be taken for a bubble and the lower (minus) sign for a drop. In these units, the value $r = 1$ corresponds to zero field.

The work of nucleation (see Eq. (7)) in the dimensionless units assumes the form

$$\Delta\Omega = Sr^2\left(\frac{1}{2}-\frac{r}{3}\right)\mp\frac{F}{r}. \tag{11}$$

Equations (10) and (11) can be solved parametrically. Eliminating field F from Eq. (10) and substituting the obtained expression into Eq. (11), we obtain

$$\frac{\Delta\Omega}{S} = r^2\left(\frac{5}{6}-\frac{2r}{3}\right). \tag{12}$$

This gives

$$\frac{F}{S} = r^4\left(1-\frac{1}{r}\right), \tag{13}$$

for a bubble and

$$\frac{F}{S} = r^4\left(\frac{1}{r}-1\right) \tag{14}$$

for a drop. Specifying the value of r and using Eqs. (12)–(14), we obtain universal field dependences of the work at the extremum for a bubble and a drop.

The dependences of $\Delta\Omega$ on the radius are plotted in Figs. 1a and 1b for a bubble and a drop, respectively, for various values of the field. In zero field, both dependences display a conventional peak corresponding to the critical size of the nucleus. The effects of the field on bubbles and drops are different. For bubbles, the peak is displaced towards larger sizes, but the value of work at the peak point becomes lower. For drops, a minimum at a certain small radius R_0, corresponding to a stable state of the drop, appears on the curves along with the peak. Drops having a radius smaller than R_0 grow to this size, while larger drops tend to reduce their size to this value. With increasing field, the points of maximum and minimum converge and the extremum of $\Delta\Omega$ vanishes at a certain critical value so that the function becomes monotonically decreasing. Under these conditions, drops of any size start growing. It should be noted that the values of work at the extrema increase with the field.

Figure 2 shows the field dependences of the radius and the work at the extrema. The critical radius of a bubble increases with the field. However, the work done for its formation decreases linearly. Hence the field facilitates vaporization in a superheated liquid. The previously prevailing opinion that the field suppresses vaporization was based on an incomplete analysis of energy expenditure for nucleation. Indeed, if the changes in the electric field energy are disregarded, the energy spent for creating a critical vapor bubble increases.

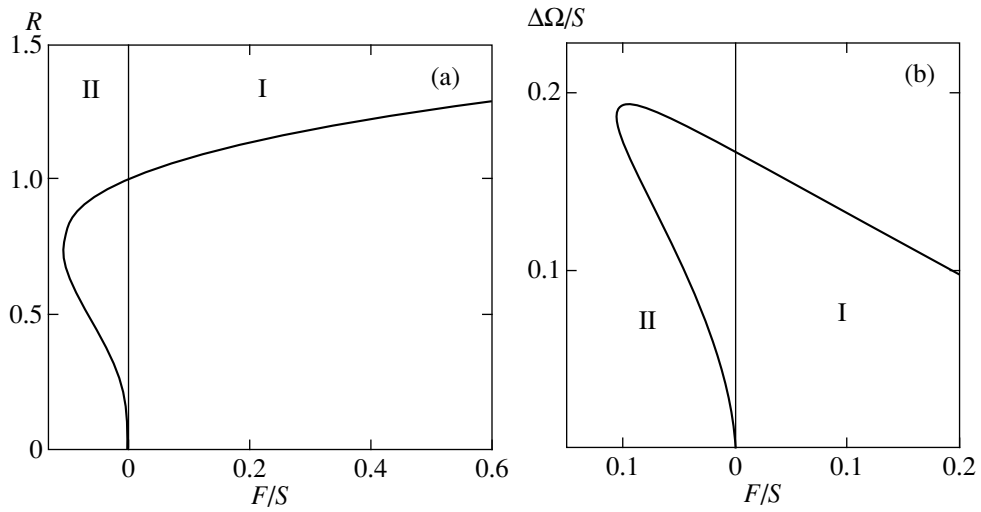

Fig. 2. Dependence of the critical radius (a) and the work of formation of a critical nucleus (b) on the field F/S of a charged center. Region I corresponds to a bubble (the field is measured from left to right) and region II corresponds to a drop (the field is measured from right to left).

In the case of drops in the presence of the field, the solution has two branches corresponding to a peak (upper branch) and a minimum (lower branch). With increasing field, these points converge and the extremum vanishes at a certain critical value of the field. The work at the extrema slightly increases with the field up to the critical value. After the attainment of this value, the formation of drops of any size becomes disadvantageous from the energy point of view. It follows from relation (9) that the value of critical field can be derived from the relation

$$F_c/S = (3/4)^3/4 = 0.105.$$

This relation corresponds to the fields

$$E[\text{MV/cm}] \approx \sqrt{0.32\,p_s[\text{atm}]}.$$

Such fields are easily attainable in experiments.

3.2. A Nucleus in a Uniform Field

A detailed analysis of the shape of a bubble (drop) as a foreign inclusion in the dielectric matrix medium in a uniform field was carried out in [6]. We consider the situation when these inclusions are another phase of the same substance. For this purpose, we must introduce some changes into the formulas derived in [6]. In a uniform field, a bubble (drop) has a shape quite close to a prolate ellipsoid of revolution in the direction of the field and having a radius R across the field and a radius $R\gamma$ ($\gamma \geq 1$) along the field. The eccentricity of such an ellipsoid can be written in the form

$$e = \sqrt{1 - 1/\gamma^2}. \tag{15}$$

The field E_1 in the ellipsoid is uniform and connected with the external field E in a uniformly unperturbed medium through the relation

$$E_1 = \varepsilon_2 E/G, \tag{16}$$

where $G = (1 - n)\varepsilon_2 + n\varepsilon_1$ and n is the depolarization factor defined as

$$n = \frac{1-e^2}{2e^3}\left[\ln\left(\frac{1+e}{1-e}\right) - 2e\right]. \tag{17}$$

Following [6], we write the balance equation for the surface forces at the pole, where only the normal component of the field differs from zero:

$$p_1 + \frac{\varepsilon_1 - \varepsilon_2}{8\pi}E_1^2\frac{\varepsilon_1}{\varepsilon_2} - \frac{2\sigma\gamma}{R} = p_2. \tag{18}$$

On the equator of the ellipsoid of revolution, the field is directed along the tangent to the interface, and the corresponding equation has the form

$$p_1 + \frac{\varepsilon_1 - \varepsilon_2}{8\pi}E_1^2 - \frac{\sigma(\gamma^2+1)}{R\gamma^2} = p_2. \tag{19}$$

Subtracting Eq. (18) from Eq. (19) and using relation (15), we obtain

$$\frac{\sigma}{R}\left(2\gamma - 1 - \frac{1}{\gamma^2}\right) = \frac{E^2(\varepsilon_1 - \varepsilon_2)^2\varepsilon_2}{8\pi G^2}. \tag{20}$$

In zero field, this equation has the obvious solution $\gamma = 1$. In the presence of the field, the nonsphericity parameter γ depends on E, σ, and R.

Let us write the expression for the work required for the formation of a nucleus in the form of an ellipsoid in a uniform field:

$$\Delta\Omega = 2\pi\sigma\left(1 + \frac{\gamma}{e}\arcsin e\right)R^2 - \frac{4\pi R^3\gamma}{3}\left(p_1 - p_0 + \frac{E^2(\varepsilon_2 - \varepsilon_1)\varepsilon_2}{8\pi G}\right). \quad (21)$$

In this equation, the last term describes the variation of the field energy upon nucleation. It should be noted that the direct evaluation of the extremum of expression (20) with respect to variables R and γ leads to expressions differing slightly from relations (19) and (20). It was demonstrated in [6], however, that this difference is insignificant.

In an analysis of expressions (19)–(21), it is convenient to go over to the dimensionless units used by us earlier. Introducing the dimensionless quantity F for a uniform field,

$$F = \frac{E^2}{8\pi p_0}, \quad (22)$$

we can write expressions (19)–(21) in the form

$$\left(2\gamma - 1 - \frac{1}{\gamma^2}\right)G^2 = \frac{2FR}{S}(\varepsilon_2 - \varepsilon_1)^2\varepsilon_2, \quad (23)$$

$$S\left(1 - \frac{\gamma}{R}\right) = \frac{F(\varepsilon_2 - \varepsilon_1)\varepsilon_1\varepsilon_2}{G^2}, \quad (24)$$

$$\Delta\Omega = SR^2\left[\frac{1}{4}\left(1 + \frac{\gamma}{e}\arcsin e\right) - \frac{R\gamma}{3}\right] - \frac{R^3\gamma}{2}\frac{F(\varepsilon_2 - \varepsilon_1)\varepsilon_2}{G}. \quad (25)$$

The system of equations (23)–(25) also has a parametric solution:

$$R = \gamma\left(1 + \frac{K(\gamma)\varepsilon_1}{\varepsilon_2 - \varepsilon_1}\right), \quad (26)$$

$$\frac{F}{S} = \frac{G^2K(\gamma)}{\left(1 + \frac{K(\gamma)\varepsilon_1}{\varepsilon_2 - \varepsilon_1}\right)(\varepsilon_2 - \varepsilon_1)^2\varepsilon_2}, \quad (27)$$

$$\frac{\Delta\Omega}{S} = \gamma^2\left(1 + \frac{K(\gamma)\varepsilon_1}{\varepsilon_2 - \varepsilon_1}\right)^2\left\{\frac{1}{4}\left(1 + \frac{\gamma}{e}\arcsin e\right) - \frac{\gamma^2}{3}\left[1 + \frac{K(\gamma)\varepsilon_1}{\varepsilon_2 - \varepsilon_1}\left(1 + \frac{G}{\varepsilon_1}\right)\right]\right\}, \quad (28)$$

where $K(\gamma) = 1 - 1/2\gamma - 1/2\gamma^3$.

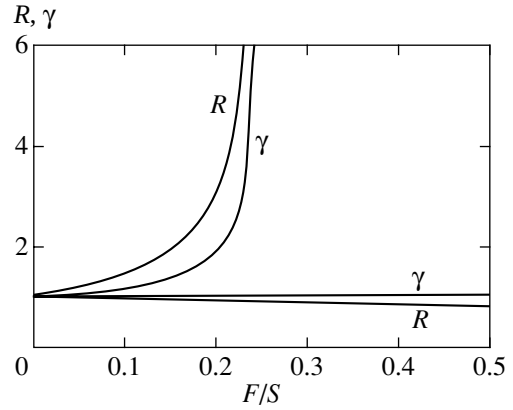

Fig. 3. Dependence of the critical radius R and parameter γ on the field F/S for a uniform field. The upper two curves correspond to a bubble ($\varepsilon_1 = 1$, $\varepsilon_2 = 3$) and the lower two curves, to a drop ($\varepsilon_1 = 3$, $\varepsilon_2 = 1$).

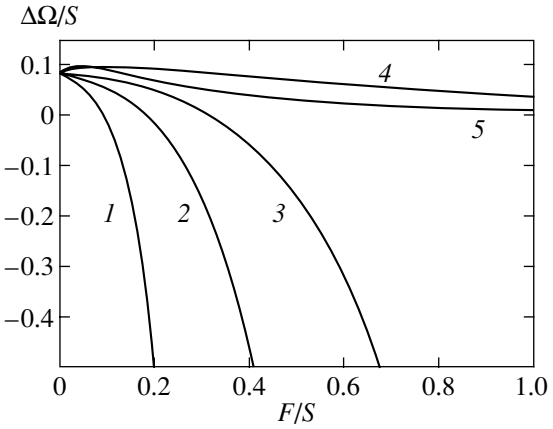

Fig. 4. Dependence of the work of formation of a critical nucleus in a uniform field on the parameter F/S. Curves *1–3* correspond to a bubble: $\varepsilon_1 = 1$, $\varepsilon_2 = 2.5$ (*1*), $\varepsilon_1 = 1$, $\varepsilon_2 = 2$ (*2*), and $\varepsilon_1 = 1$, $\varepsilon_2 = 3$ (*3*). Curves *4* and *5* correspond to a drop: $\varepsilon_1 = 3$, $\varepsilon_2 = 1$ (*4*) and $\varepsilon_1 = 8$, $\varepsilon_2 = 1$ (*5*).

Figure 3 shows the dependences of radius R and the nonsphericity parameter γ on the quantity F/S characterizing the field for a bubble and for a drop. The radius of a bubble and its elongation along the field increase as the value of the field becomes higher. The value of γ for a drop also increases. However, the radius of the drop increases.

Figure 4 shows the dependence of the work required for the formation of a nucleus of the critical size at the peak on the field F/S for a bubble ($\varepsilon_1 = 1$) for various values of the permittivity of the surrounding liquid and for a drop surrounded by vapor ($\varepsilon_2 = 1$) for various values of the permittivity of the drop itself. The work first increases and then decreases upon an increase in the field in the case of a drop. In the case of a bubble, the work decreases, this decrease becoming stronger and stronger with increasing ε_2.

Table

U, V	100	200	300
E, MV/cm	0.63	1.25	1.88
F	0.17	0.68	1.54
τ, ms (theory)	45	15.6	2.6
τ, ms (experiment)	40	10	2

4. COMPARISON WITH EXPERIMENT

The nucleation frequency (or rate) per unit volume is usually written in the following form [21]:

$$J = \exp(-G + 88), \tag{29}$$

where $G = \Delta\Omega/T$ is the Gibbs number. In experiments, the average expectation time τ for the nucleus appearance in the volume W of a metastable liquid,

$$\tau = (JW)^{-1} \tag{30}$$

is usually used.

In experiments [17, 18], an intensification of boiling of superheated liquid *n*-hexane (C_6H_{14}) in an electric field was observed. The experiments were made in a miniature bubble chamber in the temperature range from 120 to 170°C. The liquid under investigation was heated in a special heater, compressed by a buffer gas to pressures exceeding the saturation pressure at a given temperature. Then the pressure was abruptly decreased, the liquid was transformed into the superheated state, and a high-voltage pulse was supplied at this instant. The time of existence of the superheated liquid (from the beginning of the pressure drop and voltage supply to the beginning of boiling) was measured. It was found that at a fixed temperature, this time decreases considerably upon an increase of the applied voltage; i.e., the electric field initiates boiling. The time of boiling was of the order of tens of milliseconds for field strengths (0.6–5.6) MV/cm. Such fields emerged in the liquid near the edge of the inner metallic electrode and were extremely nonuniform.

Let us consider a regime of liquid superheating realized in [17, 18] when the superheating temperature was 105 K. In this case, the temperature of the liquid was $T = 447$ K, the saturation pressure was $p_s = 12$ atm, the density of the liquid was $\rho_2 = 0.476$ g/cm^3, the density of its vapor was $\rho_1 = 0.0054$ g/cm^3 (these values were calculated by us from the empirical equation for *n*-hexane from [21]), the permittivity of the liquid was $\varepsilon_2 = 1.7$, and the permittivity of vapor was $\varepsilon_1 \approx 1$. The surface tension was calculated by us from the data also presented in [21] and amounted to $\sigma = 3.9$ dyn/cm. The liquid was "unloaded" to the atmospheric pressure. Under these conditions, the degree of supersaturation introduced by us was $S = 9.75$ and the critical radius of a bubble in zero field was $R_c \approx 0.8 \times 10^{-6}$ cm. The quantity $G - 88$ can be written as $104\Delta\Omega(F) - 88$. This is an extremely sharp function of the field. For this reason, it is important to analyze the behavior of this function near zero. We expand it in the field at point F_0, which is a zero of this function. This gives

$$G - 88 = 104\alpha(F - F_0),$$

where

$$\alpha = \frac{d(\Delta\Omega)}{dF}.$$

We will estimate the value of F_0 using formula (22). This gives a dimensionless value of the field of approximately 0.7, or approximately 1 MV/cm in the dimensional units. The coefficient α for the experimental conditions is unknown. We can assume that its value is small and negative since the field under the experimental conditions is nonuniform, although it is not as strong as the field of a charged center. Finally, we can write the nucleation time in the form

$$\tau = \tau_0 \exp[104\alpha(F - 0.7)].$$

This dependence correctly describes the experimental data for $\tau_0 = 15$ μs and $\alpha = -0.02$. This is an indirect indication of the fact that the field strength decreases insignificantly over the characteristic length of the nucleus size.

The table contains the values of applied voltage U, the values of the field strengths emerging at the edge of the inner metallic electrode in dimensional and dimensionless units, and also the experimental and calculated time τ elapsed before the liquid starts boiling.

5. CONCLUSIONS

The above analysis of the possibility of formation of nuclei of a new phase in electric fields demonstrated that the electric field of a charged center initiates the condensation of a supersaturated vapor as well as the boiling of a superheated liquid. In the latter case, the size of critical nuclei is larger than that in zero field. However, the nucleation work calculated taking into account the change in the electric field energy becomes smaller. In a uniform field, the vapor condensation is also intensified. The boiling of a superheated liquid in a uniform field is facilitated and the critical size of nuclei and the work spent for their formation increase. The concepts developed here provide a qualitative explanation of the experimental results [17] if we assume that the field varies insignificantly over the sample size.

ACKNOWLEDGMENTS

The authors are grateful to I. L. Iosilevskiĭ for valuable remarks.

The work was financed by the Russian Foundation for Basic Research (project nos. 99-02-16596, 99-02-16619, and 00-15-96529).

REFERENCES

1. M. Abraham, *The Classical Theory of Electricity and Magnetism*, revised by R. Becker (Blackie, London, 1937; ONTI, Moscow, 1939).
2. J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1941; Gostekhizdat, Moscow, 1948).
3. I. E. Tamm, *Fundamentals of the Theory of Electricity* (Nauka, Moscow, 1966; Mir, Moscow, 1979).
4. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media* (Nauka, Moscow, 1992; Pergamon, New York, 1984).
5. M. A. Leontovich, *An Introduction to Thermodynamics. Statistical Physics* (Nauka, Moscow, 1983).
6. C. G. Garton and Z. Krasucki, Proc. R. Soc. London, Ser. A **280**, 383 (1964).
7. K. J. Cheng and J. B. Chaddock, Phys. Lett. A **106A**, 51 (1984).
8. A. I. Grigor'ev, A. N. Zharov, and S. O. Shiryaeva, Zh. Tekh. Fiz. **69** (8), 43 (1999) [Tech. Phys. **44**, 908 (1999)].
9. D. N. Gerasimov and O. A. Sinkevich, Teplofiz. Vys. Temp. **36**, 357 (1998).
10. B. M. Smolyak, in *Thermophysical Properties of Liquids and Explosive Boiling* (Uchebno-Nauchn. Tsentr Akad. Nauk SSSR, Sverdlovsk, 1976), p. 79.
11. A. I. Rusanov, Kolloidn. Zh. **41**, 927 (1979).
12. M. Volmer, *Kinetik der Phasenbildung* (T. Steinkopf, Dresden, 1939; Nauka, Moscow, 1986).
13. P. A. Pavlov and V. P. Skripov, in *Atomic and Molecular Physics* (Uchebno-Nauchn. Tsentr Akad. Nauk SSSR, Sverdlovsk, 1969), p. 134.
14. A. I. Zanin, E. N. Sinitsyn, and A. A. Bagrinovskiĭ, in *Thermophysical Investigations of Superheated Liquids* (Uchebno-Nauchn. Tsentr Akad. Nauk SSSR, Sverdlovsk, 1981), p. 65.
15. A. I. Zanin, in *Thermophysical Investigations of Metastable Liquids* (Uchebno-Nauchn. Tsentr Akad. Nauk SSSR, Sverdlovsk, 1986), p. 78.
16. P. L. Marston and R. E. Apfel, Phys. Lett. A **60A**, 225 (1977).
17. D. S. Parmar and A. K. Jalaluddin, J. Phys. D **6**, 1287 (1973).
18. D. S. Parmar and A. K. Jalaluddin, Phys. Lett. A **61A**, 43 (1977).
19. D. S. Parmar and B. Labroo, Phys. Lett. A **88A**, 466 (1982).
20. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics* (Nauka, Moscow, 1988; Pergamon, New York, 1987).
21. V. P. Skripov, E. N. Sinitsyn, P. A. Pavlov, G. V. Ermakov, G. N. Muratov, N. V. Bulanov, and V. G. Baidakov, *Thermophysical Properties of Liquids in Metastable State* (Nauka, Moscow, 1980).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 93, No. 4, 2001, pp. 760–770.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 871–882.

FLUIDS

Collective Effects in Doped Nematic Liquid Crystals¶

S. B. Chernyshuk[a,*], B. I. Lev[a, b], and H. Yokoyama[b]

[a]*Department of Theoretical Physics, Institute of Physics, National Academy of Sciences of Ukraine, Kiev, 03039 Ukraine*
**e-mail: chernysh@iop.kiev.ua*

[b]*Japan Science and Technology Corporation, Yokoyama Nano-Structured Liquid Crystal Project, Tsukuba Research Consortium 5-9-9, Tokodai, Tsukuba, Ibaraki 300-2635, Japan*

Received February 28, 2001

Abstract—We studied the collective elastic interaction in a system of many macroparticles embedded in a nematic liquid crystal. A theoretical approach to the interaction of macroparticles via deformation of the director field [1] is developed. It is found that the director field distortion induced by many particles leads to the screening of the elastic pair interaction potential. This screening strongly depends on the shape of the embedded particles: it exists for anisotropic particles and is absent for spherical ones. Our results are valid for the homeotropic and the planar anchoring on the particle surface and for different Frank constants. We apply our results to cylindrical particles in a nematic liquid crystal. In a system of magnetic cylindrical grains suspended in a nematic liquid crystal, the external magnetic field perpendicular to the grain orientation results in inclining the grains to the director and induces an elastic Yukawa-law attraction between the grains. The appearance of this elastic attraction can explain the cellular texture in magnetically doped liquid crystals in the presence of the magnetic field [2]. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Suspensions in liquids were recently given an intensive consideration in science and technology. Colloid suspensions of solid particles coated with a surfactant and dispersions of liquid droplets form a medium have attracted great interest in different practical applications including medicine [3, 4]. Small particles suspended in a nematic liquid crystal make a new composite material with unique physical properties that originate from the orientational ordering of the liquid crystal. Mechanical and optical properties of this medium are primarily determined by the collective behavior of these particles. Depending on their size and anchoring energies, the particles form chains [5–7], anisotropic clusters [8–10], or periodic structures [1, 11].

The origin of the structure formation is the overlapping distortions of the director field **n** caused by single particles. These distortions interfere and result in a fascinating anisotropic interaction between particles. The director deformations greatly depend on the particle sizes and anchoring energy. For the normal anchoring, the director prefers to be normal to the surface of the particles; for the planar anchoring, the director prefers to be parallel to the surface. For a strong anchoring, the boundary conditions on **n** are fixed and impose topological constrains on the director field around the particle. Topological defects arising in this case cannot be removed from the particle. A single spherical particle with a strong normal anchoring induces a point topological defect called the hyperbolic hedgehog. The droplet and the defect form a dipole that was theoretically described with the help of ansatz functions using the electrostatic analogy [5, 12]. Such dipoles play the dominant role in the formation of chains. Terentjev *et al.* [13, 14] introduced the Saturn-ring configuration with the quadrupole symmetry where a –1/2 disclination ring surrounds the sphere along the equator. It was investigated by both analytical and numerical methods. For a finite anchoring strength of the molecules at the surface, a ring configuration occurs. In this configuration, the director field is smooth everywhere, and a ring of tangentially oriented molecules is located at the equator of the sphere [14]. Using a Monte-Carlo simulation, Ruhwandl and Terentjev [15] showed that the surface ring is the preferred configuration for a sufficiently small anchoring. Stark [17] showed that the transition from the dipole to the Saturn-ring configuration can be achieved by either decreasing the particle size or applying the magnetic field or decreasing the anchoring.

In all the papers cited above, the behavior of spherical droplets in a nematic liquid crystal was studied. However, it was shown that there can be interesting properties and new structures in suspensions of anisotropic particles [1]. In 1970, Brochard and de Gennes [19] showed for the first time that the doping of a nematic liquid crystal with ferromagnetic cylindrical grains leads to a macroscopic collective behavior. This behavior is manifested as a distortion of the uniform molecular orientation of the entire matrix upon application of an external magnetic field. In other words, magnetic grains rule over the orientation of the entire nematic liquid crystal matrix. This was experimentally confirmed

¶This article was submitted by the authors in English.

by Chen and Amer [2, 20]. They found that the doped nematic exhibits a cellular texture with the cells of the order of tens of micrometers at the critical magnetic field $H \sim 30$ G. The magnetically doped nematic liquid crystal (DNLC) system in the magnetic field was theoretically examined by Burylov and Raikher [21, 22]. However, since the elastic interaction between the grains was not taken into account, the cellular texture itself has not found a satisfactory theoretical explanation. The first attempt to find the elastic interaction between cylindrical grains in a nematic matrix was made by Lopatnikov and Namiot [24], who found the pair interaction potential for the weak anchoring when the grains lie along the director.

A new approach to finding the pair interaction potential between arbitrarily shaped particles for the weak anchoring was proposed in [1]. This approach allows finding the interaction potential for any orientation of the particles with respect to the director. The pair interaction potential was found as the result of the overlapping of distortions of the director field from two particles. The general potential obtained in [1] reduces to the results of Ramaswamy *et al.* [23] and Lubensky *et al.* [12] for spherical particles and to the result of Lopatnikov and Namiot [24] for cylindrical particles. For spherical particles, this potential differs from that obtained by Ruhwandl and Terentjev only by the angular dependence of the interacting macroparticles with respect to the director.

In this paper, we use the approach proposed in [1] to consistently account for the interference of the deformations of the director field from all particles embedded in the liquid crystal. Distortions of the director field from the other particles affect the interaction potential between two chosen particles. We show that the collective screening effect arises when the concentration of particles is high. This screening effect is shown to essentially depend on the shape and orientation of the particles and on the anchoring strength. For example, the screening is absent for spherical particles and is significant for anisotropic particles, e.g., cylinders. When the cylinders are placed at an angle to the director, the screened Coulomb attraction of the Yukawa form arising between them can lead to nontrivial consequences. In this paper, we show that this potential is responsible for the cellular texture in ferronematics that was observed by Chen and Amer [2]. We show that the effective charge in the screened Coulomb attraction greatly depends on the angle between the grains and the director. It is zero in the equilibrium states when the grains are parallel or perpendicular to the director in the case of the planar or homeotropic anchoring. The external magnetic field that is not parallel to the initial orientation of the magnetic grains takes them out of the equilibrium state; the effective charge then arises because of the angle between the grains and the director. We also show that the screening is not always exponential but can be trigonometrical under some conditions. It can occur only in the presence of the external field when the angle between the grains and the director exceeds the critical threshold.

In Section 2, we formulate the problem of finding the elastic energy of a doped nematic liquid crystal following the approach in [1]. In Subsection 2.1, we find the director distribution resulting from the interference of the distortions induced by all particles. In Subsection 2.2, we consider the energy of these director distortions and extract the screened pair interaction potentials from it. In Subsection 2.3, we find an analytical expression for the pair interaction potential in the diagonal approximation. In Section 3, we find the elastic Yukawa attraction of the magnetic grains in the liquid crystal in the presence of the external magnetic field. We consider the system of many particles attracting in accordance with the Yukawa law and find the conditions for the clumping to occur in this system. This allows us to explain the cellular texture in ferronematics.

2. FORMULATION OF THE PROBLEM

The free energy of a nematic liquid crystal is given by

$$F_b = \frac{1}{2}\int d^3r\{K_{11}(\mathrm{div}\mathbf{n})^2 + K_{22}(\mathbf{n}\cdot\mathrm{rot}\mathbf{n})^2 + K_{33}(\mathbf{n}\times\mathrm{rot}\mathbf{n})^2\}, \quad (1)$$

where K_{ii} are the Frank elastic constants and $\mathbf{n}$ is the director. The macroparticles embedded in a liquid crystal induce deformations of the director field. The surface of these particles can be coated with different surfactants that orient the nematic molecules either normally or tangentially to the surface. The surface energy can be written as

$$F_s = \sum_p \oint ds W(\mathbf{s})(\nu(\mathbf{s})\cdot\mathbf{n}(\mathbf{s}))^2, \quad (2)$$

where $\nu(\mathbf{s})$ is the unit normal to the surface at the point $\mathbf{s}$ on the surface and $W(\mathbf{s})$ is the anchoring coefficient at this point. In the general case, $W(\mathbf{s}) > 0$ corresponds to the planar anchoring and $W(\mathbf{s}) < 0$ corresponds to the normal anchoring. The integral is taken over the entire surface of particle p. We assume that all particles have the same orientation in space (for example, with the help of the external field), but their centers of mass can move freely under the action of the resulting elastic potentials in order to minimize the total free energy of the system. The total free energy is the sum of the bulk and surface energies:

$$F = F_b + F_s. \quad (3)$$

We do not consider the distribution entropy part of the free energy, because it does not affect the director distribution and is irrelevant for finding the elastic interaction potential between particles.

We consider the case of the weak anchoring, where $Wr_0/K \ll 1$ (we imply the absolute value of W), where r_0 is the smallest size of the particle, e.g., the radius of

the sphere or the cylinder. For the homeotropic anchoring and spherical droplets, this corresponds to the surface ring configuration [13, 14]. In this case, we assume that the director distortion from the homogeneous state $\mathbf{n}_0$ is small everywhere:

$$\mathbf{n}(\mathbf{r}) = \mathbf{n}_0 + \delta\mathbf{n}(\mathbf{r}), \quad |\delta\mathbf{n}| \ll 1. \tag{4}$$

Under this assumption, the director smoothly varies from point to point and no topological defects arise in the vicinity of particles. This is a consequence of the weak anchoring strength and of the small particle size (we consider particles with the size less than 1 μm, which only slightly distort the director for the real anchoring strength [15]). We can use the Fourier representation for the director in the entire space, thereby considerably simplifying the problem.

In the Fourier representation, we have

$$\delta\mathbf{n}(\mathbf{r}) = \frac{1}{(2\pi)^3}\int d^3q \exp(-i\mathbf{q}\cdot\mathbf{r})\delta\mathbf{n}(\mathbf{q}). \tag{5}$$

Inserting (5) in bulk Frank energy (1), we obtain

$$F_b = \frac{1}{2}\frac{1}{(2\pi)^3}\int d^3q\{K_{11}|\mathbf{q}\cdot\delta\mathbf{n}(\mathbf{q})|^2 + K_{22}|[\mathbf{n}\times\mathbf{q}]\cdot\delta\mathbf{n}(\mathbf{q})|^2 + K_{33}|(\mathbf{n}\cdot\mathbf{q})\delta\mathbf{n}(\mathbf{q})|^2\}. \tag{6}$$

To simplify this expression, we choose the special basis

$$\mathbf{e}_1 = \frac{(\mathbf{q}_\perp\times\mathbf{n}_0)}{q_\perp}, \quad \mathbf{e}_2 = \frac{\mathbf{q}_\perp}{q_\perp}, \quad \mathbf{e}_3 = \mathbf{n}_0, \quad \mathbf{q}_\perp = \mathbf{n}_0\times\mathbf{q}. \tag{7}$$

We then have $\mathbf{q} = (q_\perp, 0, q_\parallel)$ and $\delta\mathbf{n} = (\delta n_1, \delta n_2, 0)$, and Eq. (1) therefore reduces to

$$F_b = \frac{1}{2}\frac{1}{(2\pi)^3}\times\sum_i\int d^3q\{K_{ii}q_\perp^2 + K_{33}q_\parallel^2\}|\delta n_i(\mathbf{q})|^2. \tag{8}$$

Because we assume that the director smoothly varies from point to point and relation (4) is true, we can consider the director to have a given value inside the volume of the particle. This assumption is valid if the total volume of the suspended particles is much less than the entire volume of the system, i.e., the volume fraction of the particles is small, $cv \ll 1$, where $c = N/V$ is the concentration and v is the volume of a particle (the "gas" approximation). For the real system [2], $c = 10^{10}$ cm^{-3}, $v \sim 10^{-15}$ cm^3, and $cv \sim 10^{-5}$, and our assumption is therefore true.

The director on the surface can therefore be expressed through the director at the center of mass $\mathbf{R}_p$ of the particle and its derivatives,

$$\delta\mathbf{n}(\mathbf{s}) = \delta\mathbf{n}(\mathbf{R}_p) + (\rho\nabla)\delta\mathbf{n}(\mathbf{R}_p) + \frac{1}{2}(\rho\nabla)^2\delta\mathbf{n}(\mathbf{R}_p),$$

where ρ is the vector drawn from the center of mass to the point **s** on the surface. The complete expression for the director on the surface through the director at the center of mass of the particle is therefore given by

$$\mathbf{n}(\mathbf{s}) = \mathbf{n}_0 + \delta\mathbf{n}(\mathbf{R}_p) + (\rho\nabla)\delta\mathbf{n}(\mathbf{R}_p) + \frac{1}{2}(\rho\nabla)^2\delta\mathbf{n}(\mathbf{R}_p). \tag{9}$$

We now fix a coordinate system (x, y, z) where the z axis is parallel to the undeformed director $\mathbf{n}_0$ and x and y are perpendicular to it. This system is firmly fixed in space. We next substitute director field (9) in the scalar product $(\mathbf{n}(\mathbf{s})\cdot\nu(\mathbf{s}))^2$ and also include the second powers of the perpendicular director deformations δn_x and δn_y. We thus write

$$(\nu(\mathbf{s})\cdot\mathbf{n}(\mathbf{s}))^2 = (\nu\cdot\mathbf{n}_0)^2 + 2(\nu\cdot\mathbf{n}_0)(\nu\cdot\delta\mathbf{n}) + 2(\nu\cdot\mathbf{n}_0)(\rho\cdot\nabla)(\nu\cdot\delta\mathbf{n}) + 2(\nu\cdot\mathbf{n}_0)(\rho\cdot\nabla)^2(\nu\cdot\delta\mathbf{n}) + (\nu\cdot\delta\mathbf{n})^2, \tag{10}$$

where

$$\nu = \nu(\mathbf{s}), \quad \delta\mathbf{n} = \delta\mathbf{n}(\mathbf{R}_p).$$

We note that this expression involves two smallness parameters. The first is the perpendicular component of the director,

$$|\delta n_x|, |\delta n_y| \sim \varepsilon, \quad \delta n_3 \sim \varepsilon^2,$$

and the second is the ratio $\varrho = \rho/l_n$ of the particle size to the average deformation length l_n of the director. In [1], the respective terms proportional to ε, ϱε, and $\varrho^2\varepsilon$ were taken into account. The expansion in ϱ is equivalent to the multipole expansion in [12]. In this paper, we also take the last term proportional to ε^2 into account. This term is not essential at the distances comparable to the average distance between particles, as we see below, and it can therefore be omitted for the systems considered in [6, 7, 12], where the concentration of dispersed particles is small. It becomes essential for dense colloids, where there are too many particles and where the interference of the distortions from all particles is considerable. In [12], $|\delta n_\mu|$ (with $\mu = x, y$) was shown to fall off as R^{-2} and R^{-3} depending of the dipole or quadrupole symmetry. We thus conclude that $\varepsilon \sim \varrho^2$ for the third term, which has the dipole symmetry (and $\varepsilon^2 \ll \varrho\varepsilon$ in this case), and $\varepsilon \sim \varrho^3$ ($\varepsilon^2 \ll \varrho^2\varepsilon$) for the fourth term with the quadrupole symmetry. In any case, taking the last term into account gives only small corrections at the average distances and for a small number of parti-

cles. As we see below, this is essential in the collective effect of the screening at large distances, where the concentration of particles is high. We specifically clarify this problem in what follows.

For this, we write the scalar products $(\nu(\mathbf{s}) \cdot \mathbf{n}(\mathbf{s}))$ in the local basis $(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$ associated with each particle. For example,

$$\nu(\mathbf{s}) \cdot \mathbf{n}(\mathbf{s}) = \sum_{l=1,2,3} (\nu \cdot \mathbf{k}_l)(\delta\mathbf{n} \cdot \mathbf{k}_l) = \nu_l \delta n_\mu k_{l_\mu} + \nu_l \delta n_3 k_{l_3},$$

where

$$\nu_l = (\nu \cdot \mathbf{k}_l), \quad k_{l_3} = \mathbf{k}_l \cdot \mathbf{n}_0,$$

$$\delta n_3 = -\frac{1}{2}(\delta n_x^2 + \delta n_y^2).$$

The surface energy is then written as

$$F_s = F_s^{(0)} + F_s^{(1)} + F_s^{(2)}, \tag{11}$$

$$F_s^{(0)} = N\oint ds W(\mathbf{s})(\nu(\mathbf{s}) \cdot \mathbf{n}_0)^2, \tag{12}$$

$$F_s^{(1)} = \sum_p \oint ds W(s)\nu_l \nu_m k_{l_3} \times \{2\delta n_\mu k_{l_\mu} + 2(\rho\nabla)\delta n_\mu k_{l_\mu} + (\rho\nabla)^2 \delta n_\mu k_{l_\mu}\}, \tag{13}$$

$$F_s^{(2)} = \sum_p \oint ds W(s)\nu_l \nu_m \times [\delta n_\mu \delta n_\nu k_{l_\mu} k_{m_\nu} - (\delta n_x^2 + \delta n_y^2) k_{l_3} k_{m_3}], \tag{14}$$

where N is the total number of particles in the entire volume V of the liquid crystal matrix and the respective surface terms $F_s^{(1)}$ and $F_s^{(2)}$ are linear and quadratic in δn_μ. Following [1], we now define several tensors in the basis $(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$ that characterize the anchoring energy,

$$\alpha_{kl} = 2\oint ds W(s)\nu_k(s)\nu_l(s),$$

$$\beta_{klm} = 2\oint ds W(s)\nu_k(s)\nu_l(s)\rho_m(s), \tag{15}$$

$$\gamma_{klmn} = \oint ds W(s)\nu_k(s)\nu_l(s)\rho_m(s)\rho_n(s).$$

The elastic energy F_{el}, i.e., the energy of deformations of the director field, is then given by

$$F_{\text{el}} = F_b + F_s^{(1)} + F_s^{(2)}, \tag{16}$$

$$F_s^{(1)} = \sum_p \{\alpha_{lm} + \beta_{lms}(\mathbf{k}_s \cdot \nabla) + \gamma_{lmst}(\mathbf{k}_s \cdot \nabla)(\mathbf{k}_t \cdot \nabla)\}\delta n_\mu k_{l_\mu} k_{m_3}, \tag{17}$$

$$F_s^{(2)} = \frac{1}{2} \times \sum_p \alpha_{lm}[\delta n_\mu \delta n_\nu k_{l_\mu} k_{m_\nu} - (\delta n_x^2 + \delta n_y^2) k_{l_3} k_{m_3}]. \tag{18}$$

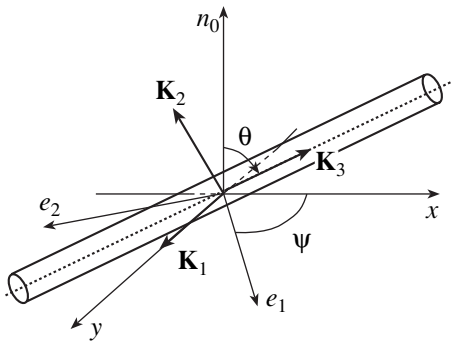

Fig. 1. Orientation of a cylindrical particle. The particle lies in the xn_0 plane at the angle θ with the director.

The main difference between this paper and [1] is in taking term (18) into account. It is quadratic in the director deformations and can be regarded as the contribution of all particles to the interference of distortions. Precisely this term leads to the screening effects. Some of its features can be considered without finding the director. For example, it is clearly seen that it vanishes for spherical particles. Indeed, $\alpha_{lm} = \alpha\delta_{lm}$ and $k_{l_\mu} k_{l_\nu} = \delta_{\mu\nu}$ for the sphere, and therefore, $F_s^{(2)} \equiv 0$. For any other shape, Eq. (18) does not vanish. To describe its effect analytically, we go to the continuum limit in this expression and replace the summation with the integration over the entire space,

$$\sum_p \longrightarrow c\int dV,$$

where, as before, $c = N/V$ is the concentration of particles:

$$F_s^{(2)} = \frac{c}{2}\int dV \tilde{a}_{\mu\nu}\delta n_\mu(\mathbf{x})\delta n_\nu(\mathbf{x}), \qquad \tilde{a}_{\mu\nu} = \alpha_{lm}[k_{l_\mu} k_{m_\nu} - k_{l_3} k_{m_3}\delta_{\mu\nu}]. \tag{19}$$

We thus consider the interference of only long-wavelength distortions of the director field. In the Fourier representation, we have

$$F_s^{(2)} = \frac{c}{2(2\pi)^3}\int d^3 q \tilde{a}_{\mu\nu}\delta n_\mu(\mathbf{q})\delta n_\nu^*(\mathbf{q}). \tag{20}$$

The tensor $\tilde{a}_{\mu\nu}$ is here taken in the (x, y, z) coordinate system, which is not convenient. It is much more suitable to write the surface energy and the bulk energy in Eq. (8) in the same basis $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$. This basis is rotated by the angle $\psi(\mathbf{q})$ with respect to (x, y, z) around the z axis (Fig. 1). In the new basis, the director has the components $\delta\mathbf{n} = (\delta n_1, \delta n_2, 0)$ and $\delta n_\mu = \varpi_{\mu i}\delta n_i$ (with $\mu = x, y$ and $i = 1, 2$). The rotation matrix is given by

$$\varpi_{\mu i} = \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix}.$$

In the basis ($\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$), the surface energy becomes

$$F_s^{(1)} = \sum_p \frac{1}{2(2\pi)^3}\int d^3q\{\exp(-i\mathbf{q}\cdot\mathbf{r}_p)a_m^*(\delta\mathbf{n}(\mathbf{q})\cdot\mathbf{k}_m) + \exp(i\mathbf{q}\cdot\mathbf{r}_p)a_m(\delta\mathbf{n}^*(\mathbf{q})\cdot\mathbf{k}_m)\}, \quad (21)$$

$$a_m = (\mathbf{k}_l\cdot\mathbf{n}_0) \times[\alpha_{lm} + i\beta_{lms}(\mathbf{q}\cdot\mathbf{k}_s) - \gamma_{lmst}(\mathbf{q}\cdot\mathbf{k}_s)(\mathbf{q}\cdot\mathbf{k}_t)],$$

$$F_s^{(2)} = \frac{c}{2(2\pi)^3}\int d^3q a_{ij}\delta n_i(\mathbf{q})\delta n_j^*(\mathbf{q}), \quad (22)$$

$$a_{ij} = \varpi_{i\mu}^T\tilde{a}_{\mu\nu}\varpi_{\nu j}. \quad (23)$$

We now add the bulk energy F_b and the surface energy F_s and find the total energy of the system

$$F_{\text{total}} = F_s^{(0)} + F_{\text{el}}$$

with the elastic energy

$$F_{\text{el}} = \frac{1}{2(2\pi)^3}\int d^3q V_{ij}(\mathbf{q})\delta n_i(\mathbf{q})\delta n_j^*(\mathbf{q}) + b_i^*(\mathbf{q})\delta n_i(\mathbf{q}) + b_i(\mathbf{q})\delta n_i^*(\mathbf{q}), \quad (24)$$

$$V_{ij}(\mathbf{q}) = (K_{ii}q_\perp^2 + K_{33}q_\parallel^2)\delta_{ij} + ca_{ij}, \quad (25)$$

$$b_i(\mathbf{q}) = \sum_p \exp(i\mathbf{q}\cdot\mathbf{r}_p)a_m k_{m_i}. \quad (26)$$

Here, $m = 1, 2, 3$; $i, j = 1, 2$; and $\delta n_i(\mathbf{q})$ and k_{m_i} are the projections of these vectors on the basis ($\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$).

2.1. Director Distribution in the Doped Nematic Liquid Crystal

Having found the complete expression for the elastic energy of a liquid crystal with particles, we can find the director at any point of the system from the extremum condition

$$\frac{\delta}{\delta n_j^*(\mathbf{q})}F_{\text{el}} = V_{ij}(\mathbf{q})\delta n_i(\mathbf{q}) + b_j(\mathbf{q}) = 0, \quad \delta n_i(\mathbf{q}) = -V_{ij}^{-1}(\mathbf{q})b_j(\mathbf{q}). \quad (27)$$

In the matrix form, the last equation becomes

$$\begin{pmatrix}\delta n_1(\mathbf{q})\\ \delta n_2(\mathbf{q})\end{pmatrix} = -\frac{1}{D}\begin{bmatrix}V_{22} & -V_{12}\\ -V_{12} & V_{11}\end{bmatrix}\begin{pmatrix}b_1(\mathbf{q})\\ b_2(\mathbf{q})\end{pmatrix}, \quad (28)$$

where

$$D = V_{11}V_{22} - V_{12}^2.$$

We can make some general conclusions from this expression concerning the behavior of the director. Distortions decrease far from the particle. However, the denominator D can vanish for some wave vectors $\mathbf{q}$, which corresponds to the appearance of oscillations of the director field. The origin of the director oscillations is purely collective, because the effect depends on the concentration $V_{12} \sim ca_{12}$. It also depends on the shape of the particles and their orientation with respect to the undeformed director $\mathbf{n}_0$. We assume that these oscillations can be observed experimentally, because they must lead to the scattering of electromagnetic waves with wave lengths commensurate to the oscillations period. Equation (28) seems to imply that

$$\lambda \sim \sqrt{K/c\alpha},$$

where α is the average value of the tensor α_{lm}; from Eq. (15), we have

$$\alpha_{lm} \sim WS,$$

where W is the anchoring energy and S is the area of the particle. The resonance wave length is therefore roughly estimated as

$$\lambda_{\text{res}} \sim \sqrt{K/WcS}. \quad (29)$$

For example, in the experiment with cylindrical grains [2], the parameters are given by $c \approx 10^{10}$ cm^{-3}, $S \approx 2\pi RL$, the radius of the grain is $R \approx 0.05$ μm, the length $L \approx 0.5$ μm, the elastic constant $K \sim 10^{-7}$ dyn, and the anchoring energy $W \sim 10^{-3}$ dyn/cm. We then find $\lambda_{\text{res}} \approx 30$ μm. Because it is always possible to vary the concentration and the anchoring energy, the resonance wave length can be in the range $\lambda \approx 10$–100 μm. In any case, this length must be larger than the average distance between the particles, $\lambda \gg \langle l\rangle$. For the experiment in [2], $\langle l\rangle \approx 4$ μm and all the assumptions are therefore valid. In the domain with the size about 30 μm, there are approximately 500 particles and their collective interaction can induce long-wavelength oscillations of the director field.

2.2. Elastic Energy and the Pair Interaction Potential between Particles

Having found the director field, we insert Eq. (28) in (24) and obtain the elastic energy of the director deformations in the DNLC:

$$F_{\text{el}} = -\frac{1}{2(2\pi)^3}\int d^3q V_{ij}^{-1}(\mathbf{q})b_i^*(\mathbf{q})b_j(\mathbf{q}) < 0. \quad (30)$$

The negative sign implies that the total free energy

$$F = F_s^{(0)} + F_{\text{el}}$$

evaluated for solution (28) is less than the energy $F = F_s^{(0)}$ for the undeformed director field $\mathbf{n}_0$. The total energy F_{el} can be represented as the sum of the pair

potentials between two particles. Indeed, we introduce the operator $\hat{A}_m$ such that

$$\hat{A}_m e^{i\mathbf{q}\cdot\mathbf{r}} = a_m e^{i\mathbf{q}\cdot\mathbf{r}}, \tag{31}$$

$$\begin{aligned}\hat{A}_m = (\mathbf{k}_l\cdot\mathbf{n}_0)[\alpha_{lm} + \beta_{lms}(\mathbf{k}_s\cdot\nabla)\\ + \gamma_{lmst}(\mathbf{k}_s\cdot\nabla)(\mathbf{k}_t\cdot\nabla)].\end{aligned} \tag{32}$$

The elastic energy F_{el} then takes the form

$$F_{\text{el}} = \frac{1}{2}\sum_{p,\,p'} U_{pp'}, \tag{33}$$

$$\begin{aligned}U_{pp'} = -\frac{1}{(2\pi)^3}\hat{A}_m^p\hat{A}_{m'}^{p'}\\ \times\int d^3q\exp[i\mathbf{q}(\mathbf{r}_p-\mathbf{r}_{p'})]V_{ij}^{-1}(\mathbf{q})k_{m_i}k_{m'_j}.\end{aligned} \tag{34}$$

The expression $U_{pp'}$ has the meaning of the pair interaction potential between particles p and p' that is caused by long-range deformations of the director field. The subscript p indicates that we must substitute

$$\nabla = \frac{\partial}{\partial\mathbf{r}_p}$$

in the operator $\hat{A}_m^p$. This expression is valid for particles of the ordinary shape and orientation. It accounts for screening effects that arise from the interference of the director field distortions by all particles.

2.3. Pair Potential in the Diagonal Approximation. Analytical Results

Although expression (34) is exact, it is too difficult to find the results analytically. In the most general case, the pair potential $U(\mathbf{R}, \mathbf{\Omega})$ depends on all the three components of the radius vector $\mathbf{R} = \mathbf{r}_p - \mathbf{r}_{p'}$, and on three Euler angles $\mathbf{\Omega}$ that determine the orientation of particles in space (we assume that all particles are oriented in the same way, and all of them therefore have the same Euler angles). To take the screening effects into account analytically, we consider particles with the rotational symmetry around one axis. For such particles, the pair potential $U(\mathbf{R}, \theta)$ depends on the angle θ between this symmetry axis and the director. If $\theta = 0$, all particles are parallel to the director. In this case, the entire DNLC has the rotational symmetry around the director $\mathbf{n}_0$ and the pair potential $U(R_\perp, R_\parallel)$ depends on the perpendicular and parallel projections of $\mathbf{R}$ with respect to $\mathbf{n}_0$. But in the case where $\theta \neq 0$, the second preferential direction arises in the system, the direction along which all the particles lie. We project this direction on the plane perpendicular to the director and let $\mathbf{s}$ denote the projection,

$$\mathbf{s}\cdot\mathbf{n}_0 = 0.$$

We then have the potential $U = U(R_\perp, R_\parallel, \varphi, \theta)$, where φ is the azimuthal angle between $\mathbf{R}$ and $\mathbf{s}$.

To obtain analytical results, we average over the angle φ. For this purpose, we average the tensor a_{ij} in Eq. (23) over the angle ψ and drop the off-diagonal terms; we call this the diagonal approximation:

$$\begin{aligned}a_{ij}(\psi,\theta) \longrightarrow \langle a_{ij}\rangle_\psi\\ = \frac{1}{2}\begin{bmatrix}\tilde{a}_{11}+\tilde{a}_{22} & 0\\ 0 & \tilde{a}_{11}+\tilde{a}_{22}\end{bmatrix} = a(\theta)\delta_{ij},\end{aligned} \tag{35}$$

where

$$a(\theta) = \frac{1}{2}(\tilde{a}_{11}+\tilde{a}_{22}).$$

This approximation makes the propagator $V_{ij}^{-1}(\mathbf{q})$ diagonal and allows us to take all the integrals analytically. The diagonal approximation is exact only for $\theta = 0$, when the entire system has the rotational symmetry in the xy plane. The coefficient $a(\theta)$ depends on the shape of the particles. For example, for cylinders with $R \ll L$, we have

$$a(\theta) = \pi RLW(2-3\sin^2\theta),$$

and for flat (pancake) particles with $R \gg h$,

$$a(\theta) = 2\pi R^2 W(1-3\cos^2\theta)$$

(where θ is the angle between the normal to the pancake plane and the director). In the diagonal approximation, the propagator therefore becomes

$$V_{ij}^{-1}(\mathbf{q}) = (K_{ii}q_\perp^2 + K_{33}q_\parallel^2 + ca(\theta))^{-1}\delta_{ij} \tag{36}$$

and the pair potential is given by

$$U_{pp'} = -\frac{1}{(2\pi)^3}\hat{A}_l^p\hat{A}_{l'}^{p'}[I_{ll'}(\mathbf{R})], \tag{37}$$

$$I_{ll'}(\mathbf{R}) = I_{1ll'}(\mathbf{R}) + I_{2ll'}(\mathbf{R}), \tag{38}$$

$$\begin{aligned}I_{1ll'}(\mathbf{R})\\ = \int d^3q e^{i\mathbf{q}\cdot\mathbf{R}}\frac{(\mathbf{k}_l\cdot[\mathbf{q}_\perp\times\mathbf{n}_0])(\mathbf{k}_{l'}\cdot[\mathbf{q}_\perp\times\mathbf{n}_0])}{q_\perp^2(K_{11}q_\perp^2+K_{33}q_\parallel^2+ca(\theta))},\end{aligned} \tag{39}$$

$$\begin{aligned}I_{2ll'}(\mathbf{R})\\ = \int d^3q e^{i\mathbf{q}\cdot\mathbf{R}}\frac{(\mathbf{k}_l\cdot\mathbf{q}_\perp)(\mathbf{k}_{l'}\cdot\mathbf{q}_\perp)}{q_\perp^2(K_{22}q_\perp^2+K_{33}q_\parallel^2+ca(\theta))}.\end{aligned} \tag{40}$$

It is easy to integrate over $\mathbf{q}$ in Eqs. (39) and (40) using the coordinate system with the basis

$$\begin{aligned}\mathbf{r}_1 = \frac{\mathbf{R}_\perp\times\mathbf{n}_0}{R_\perp},\quad \mathbf{r}_2 = \frac{\mathbf{R}_\perp}{R_\perp},\\ \mathbf{r}_3 = \mathbf{n}_0,\quad \mathbf{R}_\perp = \mathbf{n}_0\times\mathbf{R}.\end{aligned} \tag{41}$$

This basis is rotated with respect to the one in (7) by some angle φ about the axis $\mathbf{n}_0$. The quantities $q_\perp$ and $q_\parallel$ are similar in both bases. We therefore have

$$\exp(-i\mathbf{q}\cdot\mathbf{R}) = \exp\{-i[q_\perp R_\perp\cos\varphi + q_\parallel R_\parallel]\}$$

and the denominators of the fractions involved in (39) and (40) do not depend on the angle φ. Integrating over φ, we obtain

$$I_{\mu ll'}(\mathbf{R}) = \pi\int_0^\infty dq_\perp q_\perp \times\{Q^+_{l,l'}J_0(q_\perp R_\perp) + (-1)^\mu Q^-_{l,l'}J_2(q_\perp R_\perp)\} \times\int_{-\infty}^{\infty} dq_\parallel \frac{\exp(-iq_\parallel R_\parallel)}{(K_{\mu\mu}q_\perp^2 + K_{33}q_\parallel^2 + ca(\theta))}, \quad (42)$$

where μ = 1, 2 and

$$Q^{(\pm)}_{l,l'} = (\mathbf{r}_1\cdot\mathbf{k}_l)(\mathbf{r}_1\cdot\mathbf{k}_{l'}) \pm (\mathbf{r}_2\cdot\mathbf{k}_l)(\mathbf{r}_2\cdot\mathbf{k}_{l'})$$

and J_0 and J_2 are the Bessel functions.

In order to calculate these integrals, we must thoroughly scrutinize the function $a(\theta)$. As mentioned above,

$$a(\theta) = \pi RLW(2 - 3\sin^2\theta)$$

for cylindrical grains. The case where $W > 0$ corresponds to the planar anchoring, and $W < 0$ corresponds to the normal anchoring. For the planar anchoring, the equilibrium state of the grains is θ = 0 and

$$a_{\text{planar}}(0) = 2\pi RLW > 0,$$

and for the normal anchoring, the equilibrium state is θ = π/2 and

$$a_{\text{normal}}(\pi/2) = -\pi RLW > 0.$$

We thus see that $a(\theta) > 0$ in the equilibrium state independently of the anchoring type. But if the grains have a magnetic moment, the external magnetic field can lead the grains into the states where $a(\theta) < 0$. This occurs if

$$\arcsin\sqrt{2/3} < \theta < \pi/2$$

for the planar anchoring and if

$$0 < \theta < \arcsin\sqrt{2/3}$$

for the homeotropic anchoring. These states exist only because of the magnetic field. Both cases must therefore be considered. We first consider the case where $a(\theta) > 0$, which corresponds to the equilibrium states or the case of weak external fields. We write $I^{\text{exp}}_{\mu ll'}(\mathbf{R})$ for $I_{\mu ll'}(\mathbf{R})$ in this case. We introduce

$$p_\mu = \sqrt{\frac{K_{\mu\mu}}{K_{33}}}R_\parallel, \quad s = R_\perp, \quad z_\mu = \sqrt{\frac{ca(\theta)}{K_{\mu\mu}}}.$$

After the integration over the $q_\parallel$, $I^{\text{exp}}_{\mu ll'}(\mathbf{R})$ becomes

$$I^{\text{exp}}_{\mu ll'}(\mathbf{R}) = \frac{\pi^2}{\sqrt{K_{\mu\mu}K_{33}}}\int_0^\infty dq_\perp q_\perp \frac{\exp(-p_\mu\sqrt{q_\perp^2 + z_\mu^2})}{\sqrt{q_\perp^2 + z_\mu^2}} \times\{Q^+_{l,l'}J_0(sq_\perp) + (-1)^\mu Q^-_{l,l'}J_2(sq_\perp)\}. \quad (43)$$

For the Bessel functions, we have the relation

$$2\nu J_\nu(x) = xJ_{\nu+1}(x) + xJ_{\nu-1}(x),$$

which for ν = 1 gives

$$J_2(x) = \frac{2}{x}J_1(x) - J_0(x).$$

The corresponding integrals involving $J_1(x)$ and $J_0(x)$ are given by

$$\int_0^\infty dq_\perp q_\perp \frac{e^{-p_\mu\sqrt{q_\perp^2 + z_\mu^2}}}{\sqrt{q_\perp^2 + z_\mu^2}}J_0(sq_\perp) = \frac{e^{-z_\mu\sqrt{p_\mu^2 + s^2}}}{\sqrt{p_\mu^2 + s^2}}, \quad (44)$$

$$\int_0^\infty dq_\perp \frac{e^{-p_\mu\sqrt{q_\perp^2 + z_\mu^2}}}{\sqrt{q_\perp^2 + z_\mu^2}}J_1(sq_\perp) = \frac{1}{sz_\mu}\left[e^{-p_\mu z_\mu} - e^{-z_\mu\sqrt{p_\mu^2 + s^2}}\right]. \quad (45)$$

Using these relations, we find

$$I^{\text{exp}}_{\mu ll'}(\mathbf{R}) = \frac{\pi^2}{\sqrt{K_{\mu\mu}K_{33}}}\{[Q^+_{l,l'} + (-1)^{\mu+1}Q^-_{l,l'}] \times\frac{\exp(-z_\mu\sqrt{p_\mu^2 + s^2})}{\sqrt{p_\mu^2 + s^2}} + (-1)^\mu Q^-_{l,l'}\frac{2}{s^2 z_\mu} \times[\exp(-p_\mu z_\mu) - \exp(-z_\mu\sqrt{p_\mu^2 + s^2})]\}. \quad (46)$$

The pair interaction potential is then given by

$$U_{pp'} = -\frac{1}{(2\pi)^3} \times\hat{A}_l^p\hat{A}_{l'}^{p'}[I^{\text{exp}}_{1ll'}(\mathbf{r}_p - \mathbf{r}_{p'}) + I^{\text{exp}}_{2ll'}(\mathbf{r}_p - \mathbf{r}_{p'})]. \quad (47)$$

This is the potential of the elastic interaction between any particles in the diagonal approximation. It depends on the three components of the radius-vector $\mathbf{R} = \mathbf{r}_p - \mathbf{r}_{p'}$ between particles.

In the one-constant approximation where $K_{\mu\mu} = K_{33} = K$, the potential depends only on the scalar of the vector **R**,

$$U_{pp'} = -\frac{Q^+_{l,l'}}{4\pi K}\hat{A}_l^p\hat{A}_{l'}^{p'}\left[\frac{\exp(-\xi|\mathbf{r}_p - \mathbf{r}_{p'}|)}{|\mathbf{r}_p - \mathbf{r}_{p'}|}\right], \quad (48)$$

$$\xi^{-1}(\theta) = \sqrt{\frac{K}{ca(\theta)}}. \quad (49)$$

It is clearly seen that collective distortions of the director lead to the screening of the pair interaction potential with the screening length

$$\xi^{-1} \approx \sqrt{K/WcS}$$

(where W is the absolute value of the anchoring energy and S is the area of the particle). This screening occurs both for the homeotropic and for the planar anchoring. Because the concentration is only involved in the inverse screening length ξ, the limit as $c \longrightarrow 0$ gives $\xi = 0$ and brings us back to the unscreened result of Lev and Tomchuk [1], which is equivalent to the result of Lopatnikov and Namiot [24] for asymmetric cylinders.

All this is true only if $\xi^{-1} \gg \langle l \rangle$, where $\langle l \rangle = 1/\sqrt[3]{c}$ is the average distance between particles. We thus write the condition on the anchoring strength under which our approach is applicable:

$$W \ll \frac{K}{\sqrt[3]{c}S}. \quad (50)$$

2.3.1. Field-induced trigonometric screening. If the grains have the magnetic or electric moment, the external magnetic or electric field can change the angle θ between them and the director, which can result in $a(\theta) < 0$. To find the potential in this case, we must replace

$$z_\mu \longrightarrow \pm i z_\mu$$

in (46) and take half the sum of the two expressions,

$$I^{\text{trig}}_{\mu ll'}(\mathbf{R}) = \frac{1}{2}[I^{\text{exp}}_{\mu ll'}(iz_\mu, \mathbf{R}) + I^{\text{exp}}_{\mu ll'}(-iz_\mu, \mathbf{R})], \quad (51)$$

which gives

$$I^{\text{trig}}_{\mu ll'}(\mathbf{R}) = \frac{\pi^2}{\sqrt{K_{\mu\mu}K_{33}}}\{[Q^+_{l,l'} + (-1)^{\mu+1}Q^-_{l,l'}] \times \frac{\cos(z_\mu\sqrt{p_\mu^2 + s^2})}{\sqrt{p_\mu^2 + s^2}} + (-1)^{\mu+1}Q^-_{l,l'} \times \frac{2}{s^2 z_\mu}[\sin(p_\mu z_\mu) - \sin(z_\mu\sqrt{p_\mu^2 + s^2})]\}. \quad (52)$$

The pair interaction potential then takes the form

$$U_{pp'} = -\frac{1}{(2\pi)^3} \times \hat{A}^p_l \hat{A}^{p'}_{l'}[I^{\text{trig}}_{1ll'}(\mathbf{r}_p - \mathbf{r}_{p'}) + I^{\text{trig}}_{2ll'}(\mathbf{r}_p - \mathbf{r}_{p'})]. \quad (53)$$

In the one-constant approximation, this becomes

$$U_{pp'} = -\frac{Q^+_{l,l'}}{4\pi K}\hat{A}^p_l \hat{A}^{p'}_{l'}\left[\frac{\cos(\xi|\mathbf{r}_p - \mathbf{r}_{p'}|)}{|\mathbf{r}_p - \mathbf{r}_{p'}|}\right], \quad (54)$$

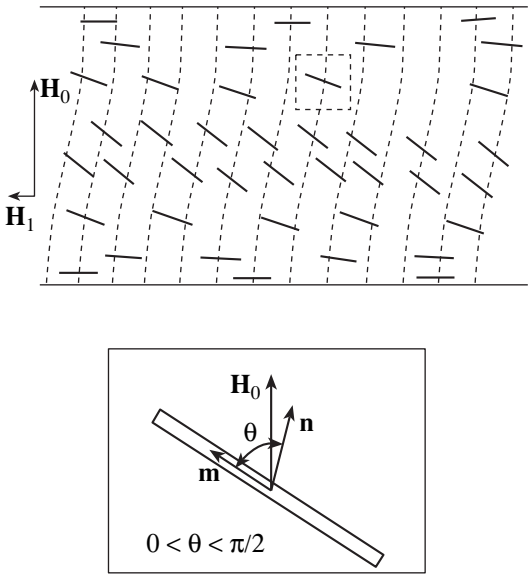

Fig. 2. Aggregation of magnetic grains in a ferronematic upon application of the magnetic field.

where

$$\xi^{-1}(\theta) = \sqrt{\frac{K}{c|a(\theta)|}}.$$

The screening becomes trigonometrical. We have obtained this result in the diagonal approximation after averaging over the azimuthal angle φ. Beyond the diagonal approximation, the screening length $\xi^{-1}(\varphi, \theta)$ actually depends on the azimuth, and the exponential screening $\xi^{-1}(\varphi, \theta_c)$ is therefore different in different directions. Changing the external field changes the angle θ, and at a certain critical angle θ_c, the screening length $\xi^{-1}(\varphi, \theta_c)$, can become infinite in some directions determined by φ; the screening thus vanishes along these directions. Subsequently increasing the field makes the screening trigonometrical along these directions. The screening is therefore exponential along certain directions and is trigonometrical along others, but it is absent on the intersections.

3. EXPLANATION OF THE CELLULAR TEXTURE IN FERRONEMATICS

In 1970, Brochard and de Gennes [19] proposed doping the liquid crystal matrix with ferromagnetic grains to allow the coupling of the liquid crystal molecular orientation to weak external fields. The authors treated such a system theoretically and predicted the Fréedericks effect in weak magnetic fields $H \sim 10$ G. The doped matrix therefore exhibits a collective orientational distortion in weak magnetic fields. They also predicted segregation effects, i.e., a smooth change of the grain concentration $c(\mathbf{R})$ from point to point in the magnetic field. In [2], the authors experimentally observed

the collective behavior in the MBBA doped with magnetic grains, which is exhibited as a long-range uniform distortion of the molecular orientation of the entire sample upon application of a weak magnetic field $H < 1$ G. In that experiment, the grains were coated with DMOAP, which provides homeotropic anchoring on their surfaces, thereby making the magnetic grains lie perpendicular to the nematic director in the absence of the magnetic field.

This system was theoretically studied by Burylov and Raikher [21, 22]. It was shown that under applying the magnetic field H, there is an angle between the grain dipole moment direction $\mathbf{m}$ (which is the unit vector along the grain) and the director $\mathbf{n}_0$; the angle is different from $\pi/2$ or 0 for a finite anchoring, as shown in Fig. 2.

To describe the experimental results on the dependence of the field-induced birefringence on the strength of the applied field, on the concentration of the magnetic dopant, and on the thickness of the nematic cell, Burylov and Raikher proposed the free energy density functional

$$F = \frac{1}{2}[K_{11}(\mathrm{div}\mathbf{n})^2 + K_{22}(\mathbf{n}\cdot\mathrm{rot}\mathbf{n})^2 + K_{33}[\mathbf{n}\times\mathrm{rot}\mathbf{n}]^2] - M_s f(\mathbf{m}\cdot\mathbf{H}) + \frac{f k_b T \ln f}{v} - \frac{fWA(\mathbf{n}\cdot\mathbf{m})^2}{d}, \quad (55)$$

where $f = cv$ is the volume fraction occupied by the particles, v is the particle volume, M_s is the magnetization inside the grains, d is the diameter, and $A \sim 1$ is a constant. This functional differs from the one proposed by Brochard and de Gennes only by the last term. The last term accounts for the weak anchoring under which $0 < \theta < \pi/2$. Minimization with respect to f (keeping the number of particles fixed) leads to

$$f = f_0 \exp\left[\frac{\mu(\mathbf{m}\cdot\mathbf{H})}{k_b T} + \frac{WAv(\mathbf{n}\cdot\mathbf{m})^2}{dk_b T}\right], \quad (56)$$

where f_0 is found from the total number of grains

$$\bar{f} = Nv = \int f(r)dV.$$

It was found that the particles accumulate in the center of the cell under applying the magnetic field (Fig. 2). For weak fields $H < 10$ G, the dependence $f(z)$ (where z is the axis perpendicular to the cell, with $z = 0$ in the center) is given by [21]

$$f(z) = \bar{f}\left[1 + \frac{\rho^2 D^2(1 - 12z^2/D^2)}{48\lambda^2}\right], \quad (57)$$

where

$$\lambda = \left(\frac{K_{33}v}{2\bar{f}k_b T}\right)^{1/2},$$

D is the thickness of the cell ($D \geq 100$ μm),

$$\rho = M_s v H/k_b T, \quad M_s \sim 340 \text{ G},$$

and $v \approx 2 \times 10^{-15}$ cm^3. At higher fields, the concentration in the center is increased faster, which was proved by computer simulations. But on reaching the field $H \sim 30$ G, experiment shows [2] that the uniform orientational distortion is replaced by a new field-induced cellular texture with the cells having dimensions on the order of tens micrometers. At the critical concentration in the center, magnetic particles clump into aggregates. This clumping had no explanation, because the magnetic dipole–dipole interaction is much smaller than the interaction with the external magnetic field. Indeed, the magnetic moment $\mu = M_s v$ induces the interaction

$$E_{dd} = \mu^2/R^3,$$

where R is the average distance between particles, $R^{-3} \sim c \sim 10^{10}$ cm^3, and therefore, $E_{dd} \sim 4 \times 10^{-15}$ erg. The energy of the interaction with the external magnetic field $H \sim 10$ G is $E_H = \mu H \sim 3 \times 10^{-12}$ erg, and hence, $E_H \gg E_{dd}$.

We explain this field-induced cellular texture by the clumping of the grains caused by elastic deformations of the director, i.e., by the elastic interaction between particles. In the one-constant approximation, this potential is given by Eq. (48). In the operators $\hat{A}_l^p$ in (32), we keep only the first term

$$\hat{A}_l = \alpha_{lm}(\mathbf{k}_l\cdot\mathbf{n}_0),$$

because the other terms give higher powers in $1/R$. For the cylinder, the tensor

$$\alpha_{lm} = 2\oint ds W(s)\nu_l(s)\nu_m(s)$$

has the components

$$\alpha_{11} = \alpha_{22} = dL\pi W, \quad \alpha_{33} = d^2\pi W,$$

and $\alpha_{lm} = 0$ for the others. Hence,

$$\alpha_{33}/\alpha_{11} = d/L \approx 0.1,$$

and we can neglect α_{33}. We thus obtain

$$U_{\mathrm{cyl}}(R) = -\frac{\alpha_{11}^2 \sin^2\theta\cos^2\theta}{4\pi K}\frac{\exp(-\xi(\theta)R)}{R}. \quad (58)$$

Cylindrical grains therefore attract each other in accordance with the Yukawa law if $\theta \neq 0, \pi/2$, which is possible in the inclined external field. In the absence of the field, the equilibrium orientations are $\theta \neq 0, \pi/2$ (depending on the planar or normal anchoring [25]) and

the potential becomes that obtained by Lopatnikov and Namiot [24], which is proportional to $1/R^3$. We set

$$e = \frac{\alpha_{11}^2 \sin^2\theta \cos^2\theta}{4\pi K}.$$

We next consider the system of particles with the concentration c and the interaction law

$$U(r) = -e\frac{\exp(-\xi r)}{r}.$$

The free energy density of this system is written as

$$F = \frac{kT}{v}\int f(\mathbf{R})\ln f(\mathbf{R})dV + \frac{1}{2v^2}\int f(\mathbf{R}) f(\mathbf{R}+\mathbf{r})U(\mathbf{r})d\mathbf{R}d\mathbf{r}. \quad (59)$$

We must find the condition for the loss of stability in this system of attracting particles. We write the concentration as

$$f(\mathbf{R}) = f_0 + \delta f(\mathbf{R}),$$

where f_0 is the ground volume fraction. Expanding

$$f(\mathbf{R}+\mathbf{r}) \approx f(\mathbf{R}) + (\mathbf{r}\cdot\nabla) f(\mathbf{R}) + \frac{1}{2}(\mathbf{r}\cdot\nabla)^2 f(\mathbf{R}),$$

we obtain

$$F - F_0 = \frac{1}{2}\int N\delta f^2(\mathbf{R}) + M(\nabla\delta f)^2, \quad (60)$$

$$N = \frac{2kT}{v} + \frac{1}{v^2}\int_{r_0}^{\infty} U(r)d\mathbf{r},$$

$$M = -\frac{1}{2v^2}\int_{r_0}^{\infty} U(r)r^2 d\mathbf{r},$$

where r_0 is the size of the particle. Inasmuch as $U < 0$, a phase transition occurs for $N < 0$. In our case, $\xi r_0 \ll 1$ and we can therefore write

$$N \approx \frac{2kT}{f_0 v} - \frac{4\pi e}{\xi^2 v^2},$$

$$M \approx \frac{12\pi e}{\xi^4 v^2}.$$

Below the critical point,

$$N \approx \frac{4\pi e}{\xi^2 v^2}.$$

The length of the first instability is

$$l_{\text{inst}} = \sqrt{\frac{2M}{N}} \approx \frac{1}{\xi}. \quad (61)$$

As discussed above, $l_{\text{inst}} \approx 30$ μm, which is in a good agreement with the experimental size of the cells [2].

4. CONCLUSION

We have derived the potential interaction for particles of the ordinary shape doped in the nematic liquid crystal. We have taken the collective screening effects into account, which is essential for the real colloid systems. It is found that the shape of the particles essentially influences the screening effects, which exist for both the homeotropic and the planar anchoring. Screening is absent for spherical particles. Anisotropic particles. (e.g., cylinders) with the magnetic or electric moment in the presence of the inclined external magnetic or electric field induce oscillations in the director distribution with the period about $\lambda \sim 10$–100 μm depending on the anchoring, the concentration, and the magnitude of the external field. In this case, selective scattering of the electromagnetic waves on these oscillations may be observed for electromagnetic waves in this range.

It is found that cylindrical grains inclined to the director attract via the Yukawa law. This explains the cellular texture in ferronematics. Application of the external magnetic field changes the orientation of the magnetic grains with respect to the director, which can lead to essentially changing the screening effects. In particular, this can lead to the trigonometric screening of the pair interaction.

Collective effects in doped nematic liquid crystals strongly depend on the anchoring strength, on the particle shape and concentration, and on external fields and make DNLC a marvellous medium for a further experimental and theoretical exploration of the different structures originating from deformations of the director field.

ACKNOWLEDGMENTS

We thank P.M. Tomchuk for very helpful discussions and H. Stark for sending us his results before publication. One of us (B.I.L.) gratefully acknowledges the financial support by the UNTC through grant no. 637.

REFERENCES

1. B. I. Lev and P. M. Tomchuk, Phys. Rev. E **59**, 591 (1999).
2. S. H. Chen and N. M. Amer, Phys. Rev. Lett. **51**, 2298 (1983).
3. D. A. Soville, W. B. Russel, and W. R. Schowaiter, *Colloidal Dispersions* (Cambridge Univ. Press, Cambridge, 1989).
4. A. P. Ruhwandl and E. M. Zukoski, Adv. Colloid Interface Sci. **30**, 153 (1989).
5. P. Poulin, H. Stark, T. C. Lubensky, and D. A. Weitz, Science **275**, 1770 (1997).
6. P. Poulin and D. A. Weitz, Phys. Rev. E **57**, 626 (1998).

7. P. Poulin, V. Cabuil, and D. A. Weitz, Phys. Rev. Lett. **79**, 4862 (1997).
8. P. Poulin, V. A. Raghunathan, P. Richetti, and D. Roux, J. Phys. I **4**, 1557 (1994).
9. V. A. Raghunathan, P. Richetti, and D. Roux, Langmuir **12**, 3789 (1996).
10. V. A. Raghunathan *et al.*, Mol. Cryst. Liq. Cryst. **288**, 181 (1996).
11. B. I. Lev, H. M. Aoki, and H. Yokoyama, submitted to Phys. Rev. E.
12. T. C. Lubensky, D. Pettey, N. Currier, and H. Stark, Phys. Rev. E **57**, 610 (1998).
13. E. M. Terentjev, Phys. Rev. E **51**, 1330 (1995).
14. O. V. Kuksenok, R. W. Ruhwandl, S. V. Shiyanovskii, and E. M. Terentjev, Phys. Rev. E **54**, 5198 (1996).
15. R. W. Ruhwandl and E. M. Terentjev, Phys. Rev. E **56**, 5561 (1997).
16. R. W. Ruhwandl and E. M. Terentjev, Phys. Rev. E **55**, 2958 (1997).
17. H. Stark, submitted to Eur. Phys. J. B.
18. H. Stark, J. Stelzer, and R. Bernhard, submitted to Eur. Phys. J. B.
19. F. Brochard and P. G. de Gennes, J. Phys. (Paris) **31**, 691 (1970).
20. B. J. Liang and S. H. Chen, Phys. Rev. A **39**, 1441 (1989).
21. S. V. Burylov and Yu. L. Raikher, Mol. Cryst. Liq. Cryst. **258**, 123 (1995); Izv. Akad. Nauk SSSR, Ser. Fiz. **55**, 1127 (1991).
22. S. V. Burylov and Yu. L. Raikher, J. Magn. Magn. Mater. **122**, 62 (1993).
23. S. Ramaswamy, R. Nityananda, V. A. Raghunathan, and J. Prost, Mol. Cryst. Liq. Cryst. **288**, 175 (1996).
24. S. L. Lopatnikov and V. A. Namiot, Zh. Éksp. Teor. Fiz. **75**, 361 (1978) [Sov. Phys. JETP **48**, 180 (1978)].
25. S. V. Burylov and Yu. L. Raikher, Phys. Rev. E **50**, 358 (1994).

Journal of Experimental and Theoretical Physics, Vol. 93, No. 4, 2001, pp. 771–776.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 883–888.

FLUIDS

Hydrodynamic Drag of a Fractal Cluster

A. L. Chernyakov

Russian Research Centre Kurchatov Institute, pl. Kurchatova 1, Moscow, 123182 Russia
e-mail: achern@lukash.asc.rssi.ru

Received April 24, 2001

Abstract—The problem of hydrodynamic drag of a fractal cluster is considered on the basis of the Brinkman equation. An analytic expression is obtained for the drag force of large and small clusters as a function of their fractal dimension. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Fractal clusters arise in different natural and technological processes. There are many papers devoted to the study of these objects [1–4]. One of important characteristics of a cluster is its hydrodynamic radius R_H, which is equal to the radius of an impermeable particle with the same hydrodynamic drag. This characteristic of a cluster is used for the experimental determination of its dimension [2, 3]. The model of a porous ball in which the distribution of the spherically symmetric density of constituent particles is a power function of the radius is frequently applied to the description of the flow of a fluid and the determination of the hydrodynamic drag of a fractal cluster. In this case, the hydrodynamics of a fluid is described by the Brinkman equation [5] with the permeability depending on coordinates. Analytical estimates are usually obtained with the use of the results of Debye and Bueche [6] for a homogeneous porous ball. Numerical methods were applied to obtain results for an arbitrary dimension of a fractal cluster [7–9]. The hydrodynamic drag of an inhomogeneous porous ball was calculated in [10] by numerical methods. In the present paper, we obtain analytic expressions for the hydrodynamic radius R_H and the fluid collection efficiency η for large and small clusters of arbitrary fractal dimension.

2. THE STUDY OF THE SOLUTIONS TO THE BRINKMAN EQUATION

The flow field $\mathbf{V}$ of a fluid around a cluster of radius R for small Reynolds numbers $\mathrm{Re} = VR/\nu \ll 1$, where ν is the kinematic viscosity, can be described by the Brinkman equation [5]

$$\Delta\mathbf{V} - \kappa^2(r)\mathbf{V} - \frac{1}{\mu}\nabla p = 0, \quad \mathrm{div}\mathbf{V} = 0, \qquad (1)$$

where $\mathbf{V}(\mathbf{r})$ is the velocity of the fluid, p is pressure, and μ is the dynamic viscosity. Suppose that $\kappa^2(r)$ is proportional to the density n of the cluster substance (sometimes, one applies more complex models [8, 9]). The number N of particles in a fractal cluster is related to its radius R by the formula

$$N = C\left(\frac{R}{a}\right)^D,$$

where a is the radius of original particles that constitute the cluster, D is the cluster dimension, and C is the proportionality factor that may depend on D. In this case, the density n is given by

$$n = \frac{CD}{4\pi a^3}\left(\frac{a}{r}\right)^{3-D}.$$

At low density, we can assume that the inverse of the permeability is proportional to the density of particles

$$\kappa^2 = \alpha 6\pi a n(r) = \kappa_0^2\left(\frac{R}{r}\right)^{3-D},$$

where the coefficient α is of the order of unity and κ_0^{-2} is the permeability on the surface of the cluster. Passing to the dimensionless variables $r' = r/R$, $\mathbf{V}' = \mathbf{V}/V_0$, and $p' = pR/\mu V_0$ and omitting the primes, we rewrite Eq. (1) as

$$\Delta\mathbf{V} - \frac{\beta^2}{r^s}\mathbf{V} - \nabla p = 0, \quad \mathrm{div}\mathbf{V} = 0, \qquad (2)$$

where

$$\beta^2 = \kappa_0^2 R^2 = \frac{2\alpha CD}{3}\left(\frac{R}{a}\right)^{D-1},$$

and $s = 3 - D$. We will seek a solution to Eq. (2) in a spherical system of coordinates with the polar axis directed along the flow incoming onto a resting cluster. Let us express the velocity of the flow in terms of the

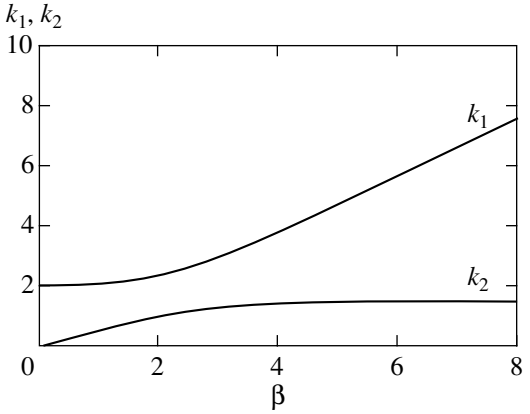

Fig. 1. Positive roots of Eq. (10) versus parameter β.

azimuthal component of the vector potential $\mathbf{V} = \mathrm{rot}\mathbf{A}_\phi$. We seek the solution in the following form:

$$\begin{aligned} p &= \cos\theta p(r), \\ A_\phi &= \frac{1}{2}\sin\theta r\chi(r), \\ V_r &= \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta A_\phi) = \cos\theta\chi(r), \\ V_\theta &= -\frac{1}{r}\frac{\partial}{\partial r}(rA_\phi) = -\frac{1}{2}\sin\theta\frac{1}{r}\frac{\partial}{\partial r}(r^2\chi(r)). \end{aligned} \tag{3}$$

Substituting these expressions into the equation for the radial component of velocity, we obtain

$$\chi'' + \frac{4}{r}\chi' - \frac{\beta^2}{r^s}\chi = p'. \tag{4}$$

Taking the divergence of the equation of motion (2), we obtain the second equation

$$p'' + \frac{2}{r}p' - \frac{2}{r^2}p = \frac{\beta^2 s}{r^{s+1}}\chi. \tag{5}$$

Equations (4) and (5) constitute the full system of equations for determining the flow field and the pressure distribution inside and outside the fractal cluster. Outside the cluster, $\beta = 0$, and the general solution to Eqs. (4) and (5) satisfying the boundary conditions at infinity is given by

$$\chi = 1 + \frac{A_1}{r} + \frac{A_2}{r^3}, \quad p = \frac{A_1}{r^2}, \tag{6}$$

where A_1 and A_2 are indefinite coefficients. Inside the cluster, the solutions to Eqs. (4) and (5) are expressed in terms of elementary functions only when $s = 0$ or $s = 2$. For a homogeneous porous ball with $s = 0$, two solutions that are regular at zero are given by

$$\begin{aligned} \chi_1 &= \frac{1}{r^{3/2}}I_{3/2}(\beta r) \\ &= \frac{1}{r^2}\left[\cosh(\beta r) - \frac{1}{\beta r}\sinh(\beta r)\right], \quad p_1 = 0, \end{aligned} \tag{7}$$

where $I_{3/2}$ is a Bessel function of a complex argument,

$$\chi_2 = 1, \quad p_2 = -\beta^2 r. \tag{8}$$

For a fractal cluster with $s = 2$, the solutions are given by

$$\chi = r^k, \quad p = \frac{2\beta^2}{k(k-1)-2}r^{k-1}. \tag{9}$$

The exponent k satisfies the fourth-order equation

$$(k^2 + 3k - \beta^2)(k^2 - k - 2) = 2\beta^2(k-1), \tag{10}$$

from which we obtain two positive roots $k_{1,2}$. The solution of this equation as a function of the parameter β is represented in Fig. 1. For small and large values of β, the roots are as follows:

$$\begin{aligned} k_1 &= 2 + \frac{1}{15}\beta^2, \quad k_2 = \frac{2}{3}\beta^2, \quad \beta^2 \ll 1, \\ k_1 &= \beta - \frac{1}{2} + \frac{9}{8\beta}, \quad k_2 = \frac{\sqrt{17}-1}{2}, \quad \beta \gg 1. \end{aligned} \tag{11}$$

To solve Eqs. (4) and (5) for an arbitrary dimension of a cluster, $0 < s < 2$, we introduce a new variable

$$y = \frac{\beta r^{1-s/2}}{1-s/2} \tag{12}$$

and a new function $\tilde{p}(r) = rp(r)$. Then, Eqs. (4) and (5) are rewritten as

$$\begin{aligned} &\frac{d^2\chi}{dy^2} + \frac{8-s}{(2-s)y}\frac{d\chi}{dy} - \chi \\ &\quad = \frac{2}{(2-s)y}\frac{d\tilde{p}}{dy} - \frac{4}{(2-s)^2y^2}\tilde{p}, \end{aligned} \tag{13}$$

$$\frac{d^2\tilde{p}}{dy^2} - \frac{s}{(2-s)y}\frac{d\tilde{p}}{dy} - \frac{8}{(2-s)^2y^2}\tilde{p} = s\chi. \tag{14}$$

One can see that, when $s \neq 2$, the points $y = 0$ and $y = \infty$ are the only singular points of this system of differential equations. It follows from the theory of linear differential equations [11] that the functions $\chi(y)$ and $\tilde{p}$ are analytic throughout the complex plane of variable y except for the points 0 and ∞. The point $y = 0$ is a regular singular point of the system of differential equations (13) and (14). In the neighborhood of zero, we can seek

solutions in the form of power series in y. We obtain the first pair of solutions by substituting the expansions

$$\chi_1(r) = r^l \sum_{n=0}^{\infty} a_n^1 \beta^{2n} r^{(2-s)n},$$
$$p_1(r) = r^{l-1} \sum_{n=0}^{\infty} b_n^1 \beta^{2n} r^{(2-s)n} \quad (15)$$

into Eqs. (4) and (5) and equating the coefficients of equal powers of r. As a result, we obtain an equation for l and a system of recurrence relations for determining the coefficients a_n^1 and b_n^1. The quantity l is determined from the quadratic equation

$$(l-1)l - 2 = 0.$$

When $l = 2$, we obtain a solution regular at zero, whereas, when $l = -1$, we obtain a singular solution. When $l = 2$, the system of equations for the unknown coefficients a_n^1 and b_n^1 is expressed as

$$[(2-s)n+2][(2-s)n+5]a_n^1 - a_{n-1}^1 = [(2-s)n+1]b_n^1,$$
$$\{[(2-s)n+1][(2-s)n+2]-2\}b_n^1 = sa_{n-1}^1. \quad (16)$$

Let us write out the solution obtained in the explicit form up to the terms proportional to β^2:

$$\chi_1(r) = r^2\left\{1 + \frac{\beta^2 r^{2-s}}{(4-s)(7-s)} \times\left[1 + \frac{s(3-s)}{(3-s)(4-s)-2}\right] + \ldots\right\}, \quad (17)$$
$$p_1(r) = r\left\{10 + \frac{\beta^2 r^{2-s} s}{(3-s)(4-s)-2} + \ldots\right\}.$$

To obtain the second independent solution, we apply the expansions of the form

$$\chi_2(r) = r^l \sum_{n=0}^{\infty} a_n^2 \beta^{2n} r^{(2-s)n},$$
$$p_2(r) = \beta^2 r^{l+1-s} \sum_{n=0}^{\infty} b_n^2 \beta^{2n} r^{(2-s)n}. \quad (18)$$

Substituting these series into Eqs. (4) and (5), we again obtain an equation for l:

$$l(l+3) = 0.$$

Only the solution with $l = 0$ is regular at zero. In this case, the system of recurrence equations is given by

$$(2-s)n[(2-s)n+3]a_n^2 - a_{n-1}^2 = [(2-s)(n-1)+1-s]b_{n-1}^2,$$
$$\{[(2-s)n+1-s](2-s)(n+1)-2\}b_n^2 = sa_n^2. \quad (19)$$

Let us write out this solution exactly up to the terms proportional to β^2:

$$\chi_2(r) = 1 + \frac{2\beta^2 r^{2-s}}{(2-s)(5-s)(3-s)} + \ldots,$$
$$p_2(r) = \beta^2 r^{1-s}\left\{\frac{1}{s-3} + \beta^2 r^{2-s} \times \frac{2s}{(2-s)(5-s)(3-s)[(3-2s)(4-2s)-2]} + \ldots\right\}. \quad (20)$$

When $\beta^2 \ll 1$, it is interesting to observe how solutions (15) and (18) are reduced to solutions (9) as $s \longrightarrow 2$. Recurrence relations (16) and (19) are easily solved in this limit:

$$a_n^1 = \frac{1}{n!}\left[\frac{1}{15(2-s)}\right]^n, \quad b_n^1 = 10a_n^1,$$
$$a_n^2 = -b_n^2 = \frac{1}{n!}\left[\frac{2}{3(2-s)}\right]^n.$$

Substituting these coefficients into series (15) and (18) and adding up the results obtained, we have

$$\chi_1(r) = r^2 \exp\left[\frac{\beta^2 r^{2-s}}{15(2-s)}\right],$$
$$p_1(r) = 10r \exp\left[\frac{\beta^2 r^{2-s}}{15(2-s)}\right],$$
$$\chi_2(r) = \exp\left[\frac{2\beta^2 r^{2-s}}{3(2-s)}\right],$$
$$p_2(r) = -\frac{1}{r}\exp\left[\frac{2\beta^2 r^{2-s}}{3(2-s)}\right].$$

One can see that, when $\beta^2 \ll 1$ and $s \longrightarrow 2$, the solutions obtained differ from the exact solutions (9) only by constant factors.

Now, let us analyze the behavior of the solutions for $y \gg 1$. The point $y = \infty$ is an irregular singular point of the system of equations (14) and (15). For one pair of solutions, $y = \infty$ is an essentially singular point, whereas, for another pair, it is a branching point. We

will seek the solutions with an essentially singular point in the form

$$\chi(r) = r^l \exp\left\{\pm\frac{\beta r^{1-s/2}}{1-s/2}\right\}\sum_{n=0}^{\infty} a_n \beta^{-n} r^{-(1-s/2)n},$$
$$p(r) = s r^{l-1} \exp\left\{\pm\frac{\beta r^{1-s/2}}{1-s/2}\right\}\sum_{n=0}^{\infty} b_n \beta^{-n} r^{-(1-s/2)n}. \quad (21)$$

Substituting these expansions into Eqs. (4) and (5), we obtain $l = -(2 - 3s/4)$ and a system of equations for the coefficients a_n and b_n. For the solution increasing at infinity, we write out the answer up to the terms proportional to β^{-1} in the following form:

$$a_0 = b_0 = 1,$$
$$a_1 = -\left[s\left(1-\frac{s}{4}\right)+\left(2-\frac{3s}{4}\right)\left(1+\frac{3s}{4}\right)\right](2-s)^{-1}, \quad (22)$$
$$b_1 = a_1 + (4-s).$$

The solution with a branching point at infinity is obtained from the expansion in inverse powers of y:

$$\chi_2 = r^{m+1}\sum_{n=1}^{\infty} a_n \beta^{-2n} r^{-(2-s)n},$$
$$p_2 = r^m \sum_{n=0}^{\infty} b_n \beta^{-2n} r^{-(2-s)n}. \quad (23)$$

Equating the terms with equal powers of r in Eqs. (4) and (5), we obtain the recurrence relations

$$[m+1-(2-s)n][m+4-(2-s)n]a_n - a_{n+1} = [m-(2-s)n]b_n,$$
$$\{[m-(2-s)n][m+1-(2-s)n]-2\}b_n = s a_{n+1}.$$

Taking into account that $a_0 = 0$ for $n = 0$, we obtain the following equation for m from this system:

$$m^2 + (1+s)m - 2 = 0.$$

It turns out that one solution increases at infinity, whereas another tends to zero. Note that the solutions that decrease at infinity must be singular at zero since the modulus of an analytic function cannot have a maximum inside the domain of analiticity [12]. Therefore, we will focus only on the solutions that increase at infinity. For the function that increases at infinity, we have

$$m = \sqrt{\left(\frac{1+s}{2}\right)^2 + 2} - \frac{1+s}{2}. \quad (24)$$

Below, we will need only the first terms of the expansion

$$\chi_2 = -m\beta^{-2} r^{m+1-(2-s)}, \quad p_2 = r^m. \quad (25)$$

3. HYDRODYNAMIC RADIUS OF A FRACTAL CLUSTER

A general solution to the system of Eqs. (4) and (5) inside the fractal cluster can be expressed as

$$\chi(r) = B_1\chi_1(r) + B_2\chi_2(r),$$
$$p(r) = B_1 p_1(r) + B_2 p_2(r),$$

where $(\chi_1(r), p_1(r))$ and $(\chi_2(r), p_2(r))$ are two independent solutions that are bounded at zero. Now, knowing the structure of the solution inside the fractal cluster, we can sew it together with the solution outside the cluster. The components of the velocity and the components of the momentum flux tensor must be continuous across the cluster surface (when $r = 1$) [13]:

$$\sigma_{rr} = -p + 2\frac{\partial V_r}{\partial r},$$
$$\sigma_{r\theta} = \frac{1}{r}\frac{\partial V_r}{\partial \theta} + \frac{\partial V_\theta}{\partial r} - \frac{V_\theta}{r}.$$

This yields the continuity of the functions χ, χ', and χ'' and the pressure p at $r = 1$. We obtain the following system of linear equations for the coefficients A_1, A_2, B_1, and B_2:

$$1 + A_1 + A_2 = B_1\chi_1 + B_2\chi_2,$$
$$-A_1 - 3A_2 = B_1\chi_1' + B_2\chi_2',$$
$$2A_1 + 12A_2 = B_1\chi_1'' + B_2\chi_2'', \quad (26)$$
$$A_1 = B_1 p_1 + B_2 p_2.$$

This system allows one to obtain all expansion coefficients from the values of the functions χ_1 and χ_2 and their derivatives at the point $r = 1$. Of greatest interest is the coefficient A_1; the drag force F of a cluster and the dimensionless hydrodynamic radius R_H are expressed in terms of this coefficient:

$$R_H = \frac{F}{6\pi\mu V_0 R} = \frac{3}{2}A_1.$$

The system of Eqs. (26) yields

$$A_1 = -\frac{3}{2}\left\{1 - \frac{3\chi_2 + \chi_2'}{2p_2} + \left(1 + \frac{\beta^2\chi_2 + p_2'}{2p_2}\right)\times\frac{p_2(3\chi_1 + \chi_1') - p_1(3\chi_2 + \chi_2')}{p_2(\beta^2\chi_1 + p_1') - p_1(\beta^2\chi_2 + p_2')}\right\}^{-1}. \quad (27)$$

Another quantity of interest is the fluid collection efficiency η, which is equal to the ratio of the fluid flow

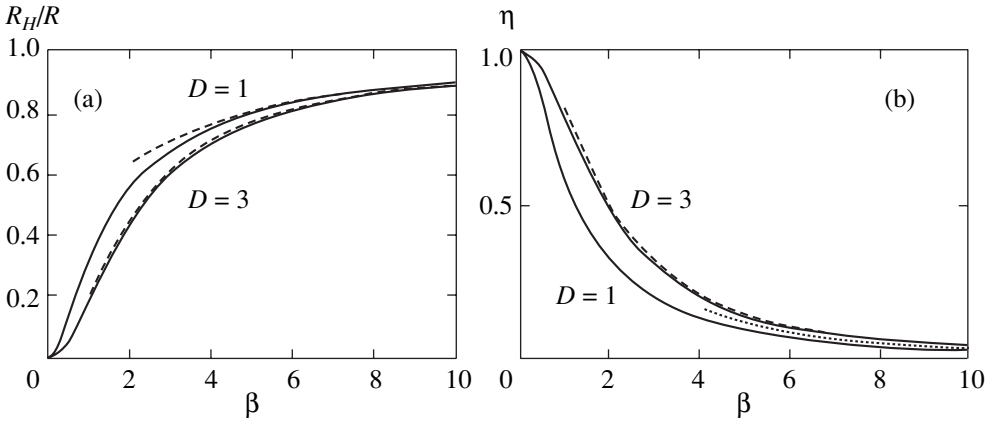

Fig. 2. (a) Hydrodynamic radius R_H and (b) fluid collection efficiency η versus β for two values of the fractal dimension D of the cluster, $D = 1$ and $D = 3$.

through a porous medium to the fluid flow at infinity that passes through an area of πR^2:

$$\eta = \frac{2}{R^2 V_0}\int_0^{\pi/2} V_r \sin\theta d\theta = B_1\chi_1 + B_2\chi_2.$$

We obtain the following expression for this coefficient:

$$\eta = -A_1\left\{\frac{\chi_2}{p_2} + \frac{(p_1\chi_2 - p_2\chi_1)\left(2 + \dfrac{\beta^2\chi_2 + p_2'}{p_2}\right)}{p_2(\beta^2\chi_1 + p_1') - p_1(\beta^2\chi_2 + p_2')}\right\}. \quad (28)$$

In the case of a homogeneous porous ball with $s = 0$, substituting exacts solutions (7) and (8) into Eq. (27), we obtain the known expression [6] for the hydrodynamic radius,

$$R_H = \left\{\frac{3}{2}\beta^{-2} + \left(1 - \frac{\tanh\beta}{\beta}\right)^{-1}\right\}^{-1} \quad (29)$$

and the coefficient η,

$$\eta = \frac{9}{2\beta^2}R_H. \quad (30)$$

For $\beta \ll 1$, from (29) we obtain

$$R_H = \frac{2}{9}\beta^2\left(1 - \frac{4}{15}\beta^2\right),$$

whereas, for $\beta \gg 1$, we have

$$R_H = \left(1 + \beta^{-1} + \frac{5}{2}\beta^{-2} + \beta^{-3}\right)^{-1}$$

exactly up to β^{-3}. In another limit case, when $s = 2$, the substitution of exact solutions (9) into Eqs. (27) and (28) yields the following expressions for the hydrodynamic radius and the fluid collection efficiency η:

$$R_H = \left\{1 - \frac{1}{4\beta^2}(k_2+3)(k_2^2 - k_2 - 2) + \frac{k_2(k_2+1)}{4\beta^2}\right.$$
$$\left.\times\frac{[(k_1+3)(k_1^2 - k_1 - 2) - (k_2+3)(k_2^2 - k_2 - 2)]}{k_1(k_1+1) - k_2(k_2+1)}\right\}^{-1},$$

$$\eta = \frac{3}{4\beta^2}R_H\left\{(2 + k_2 - k_2^2) + k_2(k_2+1)\right.$$
$$\left.\times\frac{k_1(k_1-1) - k_2(k_2-1)}{k_1(k_1+1) - k_2(k_2+1)}\right\},$$

where k_1 and k_2 are the roots of Eq. (10). For $\beta \ll 1$, we have

$$R_H = \frac{2}{3}\beta^2\left(1 - \frac{4}{3}\beta^2\right), \quad \eta = 1 - \frac{8}{9}\beta^2.$$

For large clusters with $\beta \gg 1$,

$$R_H = \left\{1 + \beta^{-1} + \frac{\sqrt{17}+3}{4}\beta^{-2} - \frac{23}{8}\beta^{-3}\right\}^{-1}, \quad (31)$$

$$\eta = \beta^{-2}\frac{3}{2}R_H(k_2+1) = \beta^{-2}R_H\frac{3(\sqrt{17}+1)}{4} \quad (32)$$

exactly up to β^{-3}.

Figure 2 represents exact values of the hydrodynamic radius R_H and the fluid collection efficiency η as functions of the parameter β for two limit values of the dimension, $D = 3$ and $D = 1$. The dashed lines represent the asymptotic functions that are exact up to β^{-3}.

Now, using the solutions obtained above for arbitrary s, we calculate the hydrodynamic radius R_H and the coefficient η for large and small values of β. When

$2\beta^2 \ll 2 - s$, the substitution of the expansions (17) and (20) into (27) and (28) yields

$$R_H = \frac{2\beta^2}{3(3-s)}\left\{1 - \frac{2\beta^2}{3(3-s)} \times\left[1 + \frac{2s^2 - 15s + 10}{(5-s)[(3-2s)(4-2s)-2]}\right]\right\} \quad (33)$$

and

$$\eta = 1 - \frac{2\beta^2}{(3-s)(5-s)} \times\left[6 - s - \frac{3s}{(3-2s)(4-2s)-2}\right]. \quad (34)$$

For $s = 0$, these expressions reduce to the formulas for a homogeneous porous ball; the first term in expansion (33) is valid for an arbitrary dimension $0 < s < 2$ and coincides with the expression obtained for $s = 2$.

For large clusters with $\beta \gg 1$, the substitution of solutions (21) and (25) into (27) and (28) yields

$$R_H = \left\{1 + \beta^{-1} + \beta^{-2}\frac{1}{4}[\sqrt{(s+1)^2 + 8} - 2s + 7] + \beta^{-3}\left[1 - \frac{s}{8}\left(17 - \frac{3}{4}s\right)\right]\right\}^{-1} \quad (35)$$

and

$$\eta = \beta^{-2}\frac{3}{2}R_H(m+2) = \beta^{-2}\frac{3}{2}R_H\left[\sqrt{\left(\frac{1+s}{2}\right)^2 + 2} - \frac{1+s}{2} + 2\right]. \quad (36)$$

One can see that the asymptotic expressions for the hydrodynamic radius R_H (35) and the fluid collection efficiency η (36) for a fractal cluster of arbitrary dimension $D = 3 - s$ exactly reduce to the expressions for $s = 0$ and $s = 2$ obtained above. As follows from (35), the effect of the dimension on the hydrodynamic radius for $\beta \gg 1$ manifests itself only starting from the terms proportional to β^{-2}. Note that the diagrams of $R_H(\beta)$ and $\eta(\beta)$ in Fig. 2, drawn for the dimensions $D = 3$ and $D = 1$, restrict the range of variation of the corresponding quantities under the variation of the dimension within the interval $3 > D > 1$.

ACKNOWLEDGMENTS

I am grateful to A.A. Kirsh and I.B. Stechkina for discussing this paper.

REFERENCES

1. B. M. Smirnov, *Physics of Fractal Clusters* (Nauka, Moscow, 1991).
2. E. P. Emets, A. É. Novoselova, and P. P. Poluéktov, Usp. Fiz. Nauk **164**, 959 (1994) [Phys. Usp. **37**, 881 (1994)].
3. S. Nyeki and I. Colbeck, J. Aerosol Sci. **25**, 75 (1994).
4. E. F. Mikhaĭlov and S. S. Vlasenko, Usp. Fiz. Nauk **165**, 263 (1995) [Phys. Usp. **38**, 253 (1995)].
5. H. C. Brinkman, Appl. Sci. Res., Sect. A **1**, 27 (1947).
6. P. Debye and A. M. Bueche, J. Chem. Phys. **16**, 573 (1948).
7. W. van Saarloos, Physica A (Amsterdam) **147**, 280 (1987).
8. S. Veerapaneni and M. R. Wiesner, J. Colloid Interface Sci. **177**, 45 (1996).
9. M. Vanni, Chem. Eng. Sci. **55**, 685 (2000).
10. G. Ooms, P. F. Mijnlieff, and H. L. Beckers, J. Chem. Phys. **53**, 4123 (1970).
11. A. Kratzer and W. Franz, *Transzendente Funktionen* (Akademische Verlagsgesellschaft, Leipzig, 1960; Inostrannaya Literatura, Moscow, 1963).
12. M. A. Lavrent'ev and B. V. Shabat, *Methods of the Theory of Functions of a Complex Variable* (Nauka, Moscow, 1973).
13. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics* (Nauka, Moscow, 1986; Pergamon, New York, 1987).

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 93, No. 4, 2001, pp. 777–785.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 889–898.

SOLIDS
Structure

Statistics of Internal Excitations of Atomic Systems¶

R. S. Berry[a] and B. M. Smirnov[b, *]

[a]*Department of Chemistry, University of Chicago, Chicago, IL 60637 USA*
[b]*Institute of High Temperatures, Moscow, 127412 Russia*
**e-mail: smirnov@orc.ru*

Received June 6, 2001

Abstract—The character of internal excitations is compared for phase transitions and chemical transitions in atomic systems. Although the temperature dependences of some physical parameters of atomic systems have resonance-like structures with maxima in both cases, the dependences of the partition functions on the number of elementary excitations or the excitation energy differ because of the difference in the numbers of interactions that govern the transitions. The phase changes of condensed rare gases are considered in the case where the external pressure is small and the differences between phases are predominantly associated with differences in configurations. Important energy parameters of rare gases are determined by the attractive part of the pairwise interaction potential between atoms. The statistical analysis shows the existence of a "freezing limit" temperature for these systems, below which the liquid state becomes unstable. The kinetics of decay of such unstable states is analyzed in terms of the diffusion of voids. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

A phase transition corresponds to a transition between different aggregate states; for a first-order phase transition, the internal energy of a bulk system changes discontinuously as the temperature varies and the pressure is constant [1–3]. In contrast to this, a chemical transition, i.e., a transition between two limiting chemical states of a substance, occurs throughout some temperature range, with a shifting equilibrium ratio of the species, when conducted at a constant pressure. This principal difference between the phase and chemical transformations is lost for systems consisting of a finite number of atoms, notably, clusters [4–10]. Computer simulations of phase transitions in clusters [4–10] reveal some peculiarities of this phenomenon, in particular, the coexistence of the phases throughout some temperature range. From general considerations, one can infer that this range has sharp upper and lower bounds, which we may call the "melting limit" T_m and the "freezing limit" T_f. The range between T_f and T_m can remain or shrink to zero as the number of particles N in the cluster grows very large. However in either case, the observable effect of increasing N is to make the range of the apparent coexistence shrink to a single temperature T_{eq} at which the free energies and mean chemical potentials of the two forms are equal. Away from this temperature, the thermodynamically unfavored phase can be present in observable amounts for relatively small N, but as N increases, the unfavored phase becomes so unfavored that the amounts or frequencies of its appearance become unobservably small. Because of the observability of unfavored phases for small systems, the transitions or phase changes between aggregate states are very similar to those between chemical states such as chemical isomers.

The study of phase transitions in clusters [4–10], especially focusing on their microscopic nature, has given us a deepened understanding of the nature of phase transitions for bulk systems. Analyzing the phase and chemical transformations in bulk systems from the microscopic standpoint, one can find both common and different features of these phenomena. Such an analysis is the goal of this paper. We are guided by the simplest cases for this analysis.

We compare the solid–liquid phase transition for a system of atoms bound by a pair interaction with the simple chemical transformations, ionization and dissociation. This comparison allows us to establish the common and different features of the phase and chemical transitions from the microscopic standpoint. The microscopic interpretation allows us to expand our understanding of the phase transition.

2. CHEMICAL EQUILIBRIUM AND TRANSFORMATIONS

We first consider the simplest chemical equilibria, in which a gaseous system consists of particles XY at low temperatures that dissociate into X and Y at high temperatures, and the chemical equilibrium therefore has the form

$$XY \longleftrightarrow X + Y. \qquad (2.1)$$

For the ionization equilibrium, we use this form and consider X to be an ion (A^+) and Y an electron (e), in

¶This article was submitted by the authors in English.

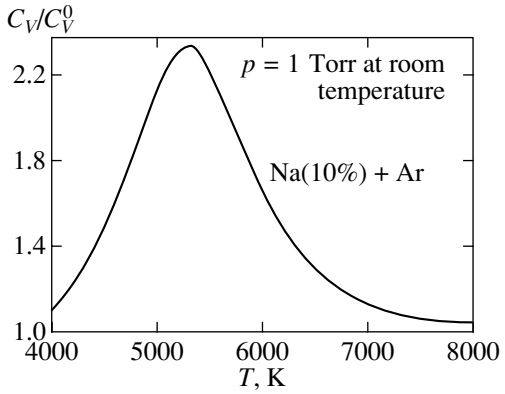

Fig. 1. The specific heat capacity of argon with a sodium admixture in the temperature range of the sodium ionization transition. The concentration of sodium atoms is equal to 10% and the number density of argon atoms corresponds to the pressure 1 Torr at room temperature.

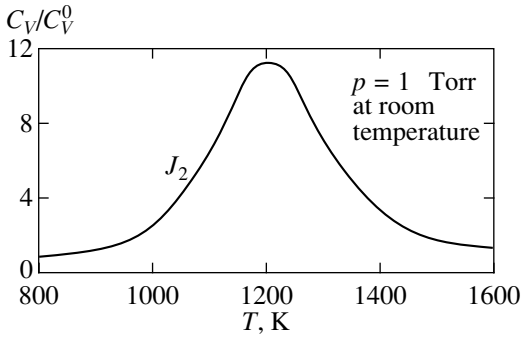

Fig. 2. The specific heat capacity of iodine J_2 in the range of the dissociation transition if the iodine number density corresponds to the pressure 1 Torr at room temperature.

which case the ionization equilibrium is

$$A^+ + e \longleftrightarrow A. \tag{2.2}$$

In parallel, we consider the dissociation equilibrium, in which the composite particles XY are dimer molecules and the dissociation equilibrium has the form

$$A + A \longleftrightarrow A_2. \tag{2.3}$$

We use the simplest formulas for these equilibria. For the ionization equilibrium, the number densities of electrons N_e, ions N_i, and atoms N_a are related by the Saha equation [11, 12]

$$\frac{N_e N_i}{N_a} = \frac{g_e g_i}{g_a}\left(\frac{m_e T_e}{2\pi\hbar^2}\right)^{3/2}\exp\left(-\frac{J}{T_e}\right), \tag{2.4}$$

where m_e and T_e are the mass and temperature of electrons and g_e, g_i, and g_a are the statistical weights of the electron ($g_e = 2$), the ion, and the atom electronic states. Introducing the probabilities for an electron to be free $w_e = N_e/N$ or to be bound in an atom $w_a = N_a/N$ (where $N = N_e + N_a$ is the total number density of free and bound electrons, and hence, $w_e + w_a = 1$), we represent Saha formula (2.3) for a quasineutral plasma with $N_e = N_i$ as

$$\frac{w_e^2}{w_a} = g\exp\left(-\frac{J}{T_e}\right), \tag{2.5}$$

where the statistical weight g of free electron states is

$$g = \frac{1}{N}\frac{g_e g_i}{g_a}\left(\frac{m_e T_e}{2\pi\hbar^2}\right)^{3/2}. \tag{2.6}$$

In general, the statistical weight of free, continuum states is the ratio of the typical atomic number density in the condensed phase to the density of free atoms in the gas phase, and hence, this value is large. Therefore, the value $1/\ln g$ is a small parameter of the theory, because we consider transitions between free and bound states; these transitions occur in a relatively narrow range of temperatures due to a small value of this parameter. We note that we deal with an ensemble consisting of a sufficiently large number of particles to allow us to neglect the fluctuations.

We now determine the temperature width for this transition. We define the electron temperature for the ionization transition T_{ion} as the temperature for which $w_e(T_{\text{ion}}) = 1/2$. To be precise, we define the temperature range ΔT of the transition from atoms to electrons and ions such that the value w_e varies in this region from 1/4 up to 3/4. Hence, we have

$$\Delta T = \frac{T_{\text{ion}}^2}{J}\ln 27, \tag{2.7}$$

and a small parameter in this case is

$$T_{\text{ion}}/J = 1/\ln g.$$

In particular, under conditions of Fig. 1, Eq. (2.7) gives

$$\Delta T/T_{\text{ion}} = 0.3.$$

Using the same expressions for the probabilities of free and bound states, we have for the dissociation equilibrium

$$\frac{w_a^2}{w_m} = g\exp\left(-\frac{D}{T_e}\right), \tag{2.8}$$

where w_a is the probability for an atom to be free and w_m is the probability for an atom to be bound in the dimer molecule; D is the dissociation energy of the molecule. In contrast to the ionization equilibrium, where the probability of excited atom states is small, excited rotational and vibrational molecule states are taken into account in Eq. (2.8). But the structure of this formula is the same as for the ionization equilibrium. Figures 1 and 2 give some examples of the ionization and dissociation equilibria. The width of the transition between free and bound states in the temperature scale

is determined by the small parameter $1/\ln g$ and decreases as the number density of particles decreases.

As a result of renormalization, we can infer from Eq. (2.5) the respective partition functions of free and bound electron states Z_e and Z_a. For the ionization equilibrium, we then have

$$\frac{Z_e^2}{Z_a} = g\exp\left(-\frac{J}{T_e}\right). \quad (2.9)$$

We obtain a corresponding relation for the dissociation equilibrium. These formulas can be used for the analysis of statistics of chemical equilibria.

We now determine the partial partition function for the ionization equilibrium that corresponds to a given number of free electrons at a certain temperature. An equivalent result appears for the dissociation equilibrium. If the total number of nuclei in the system is n and the number of ionized atoms is m, the probability of this event is determined by the Poisson formula

$$W_{nm} = C_n^m w_e^m w_a^{n-m} = C_n^m w_e^m (1-w_e)^{n-m}, \quad (2.10)$$

where the ratio between w_e and w_a (with $w_e + w_a = 1$) is given by Eq. (2.5). The partition function Z_{nm} of the system with a given number of free and bound electrons is proportional to this value, and we take these values to be identical for simplicity. We note that formation of m free electrons in this system corresponds to the excitation energy

$$m(J + 3T_e/2) \approx mJ.$$

For a large number of free and bound electrons in the system, i.e., $m \gg 1$ and $n \gg 1$, the partition function Z_{nm} has a sharp maximum as a function of m: near the maximum $m = m_0$, it has the form

$$Z_{nm} = Z_0\exp[-\alpha(m-m_0)^2], \quad (2.11)$$

and in accordance with to the above relations, we have

$$m_0 = nw_e, \quad \alpha = \frac{1}{2m_0}\frac{n}{n-m_0}. \quad (2.12)$$

From this expression, we see that the partition function has a narrow maximum in the range of the number of broken bonds $\Delta m \sim \sqrt{n}$ if $m_0 \sim n$, and the relative maximum width $\Delta m/m_0$ tends to zero as $1/\sqrt{n}$ when the number of atoms tends to infinity.

3. CONFIGURATION EXCITATION OF A SYSTEM OF BOUND ATOMS

It follows from the above analysis that as a function of the excitation energy, the partition function is characterized by one maximum. In contrast to this, the partition function for the solid–liquid phase equilibrium has two maxima, and each maximum corresponds to a certain aggregate state. Below we consider, from this standpoint, the solid–liquid phase transition for condensed rare gases when the interaction between nearest neighbors dominates. We use measured parameters of the aggregate states of condensed rare gases near their triple points and consider a condensed rare gas as a system of bound atoms with short-range interactions, i.e., assume that interactions between nearest neighbors give the main contribution to the constitutive parameters of such systems. The reduced parameters of condensed rare gases are expressed through the parameters of the pair interaction potential of these atoms, which are known well [13–16]; this information is based on the parameters of diatomic molecules, condensed and dense rare gases and collision parameters of pairs of atoms. Next, the classical character of the atomic motion in condensed systems of Ne, Ar, Kr, and Xe, together with the short-range character of the interactions of atoms in these systems, allows us to use a scaling analysis to express bulk parameters of these systems through parameters of the interaction potential of two atoms. Simplifying this operation, we choose the parameters of the pair interaction potential of these atoms as D, the depth of the potential well, and R_e, the equilibrium distance between atoms of the diatomic molecule. Adding to these parameters the atom mass m, we can express the physical value of any dimensioned property through the three parameters m, R_e, and D. The degree of coincidence of the reduced physical parameters for different rare gases then determines the accuracy of such a scaling law for real systems; for real rare gases, this measure of accuracy is several percent.

The analysis of parameters of the solid rare gases shows a small contribution of a long-range interaction. Indeed, the ratio of the distance a between nearest atoms in the solid rare gases at zero temperature to the equilibrium distance between atoms of the corresponding diatomic, averaged over all the stable rare gases, is [17–19]

$$a/R_e = 1.005 \pm 0.013,$$

and the reduced sublimation energy ε_{sub} of solid rare gases is 6.4 ± 0.2 per atom. For a system of bound atoms with only the nearest-neighbor interactions, these values are equal to 1 and 6 respectively, whereas for a system of atoms with the Lennard-Jones interactions, for which there is a long-range contribution to physical parameters of the system, the respective values are equal to 0.97 and 8.41 [20]. This shows that condensed rare gases are close to systems of atoms with the interaction between nearest neighbors only, and the error from this assumption is less than 10% for any parameter.

Taking real condensed rare gases as a system of bound atoms with a short-range interaction (i.e., the interaction between nearest neighbors only), we obtain additional information about this system on the basis of parameters of condensed rare gases. In particular, Table 1 contains the reduced parameters of the phase transition

Table 1. Reduced average parameters of condensed rare gases at the triple point [17–19]

Reduced quantity	Average value
T_{tr}/D	0.579 ± 0.007
$p_{tr}R_e^3/D$, 10^{-3}	1.9 ± 0.2
$\Delta H_{fus}/D$	0.98 ± 0.02
ΔS_{fus}	1.68 ± 0.03
$p_{tr}\Delta V/\Delta H_{fus}$, 10^{-4}	2.2 ± 0.4
ε_{sub}/D	6.4 ± 0.2
$\rho(0)R_e^3/\sqrt{2}$	1.01 ± 0.04
$\rho_{sol}R_e^3/\sqrt{2}$	0.92 ± 0.02
$\rho_{liq}R_e^3/\sqrt{2}$	0.80 ± 0.02
V_{liq}/V_0	1.153 ± 0.006
ε_{ev}/D	5.4 ± 0.2
$p_{sol}R_e^3/D$	110 ± 20
$p_{liq}R_e^3/D$	25 ± 4
ε_{liq}/D	5.5 ± 0.1
T_{cr}/D	1.04 ± 0.02

of this system near the triple point. In this table, T_{tr} and p_{tr} are the temperature and pressure at the triple point, ΔH_{fus} and ΔS_{fus} are the fusion energy and the entropy change per atom as a result of the phase transition, ρ_{sol} and ρ_{liq} are the density of the solid and liquid rare gases at the triple point, and $\rho(0)$ is this value at zero temperature; the specific volume per atom for the solid and liquid states are denoted as V_0 and V_{liq}, respectively. In addition, the equilibrium vapor pressure p above the solid and liquid surfaces is given by the formulas

$$p = p_{sol}\exp\left(-\frac{\varepsilon_{sol}}{T}\right), \quad p = p_{liq}\exp\left(-\frac{\varepsilon_{liq}}{T}\right), \qquad (3.1)$$

where the parameters ε_{sol} and ε_{liq} characterize the binding energies per atom for the solid and liquid states of this system, to within the accuracy of the thermal energy at the melting point. The parameters of Eq. (3.1) are given in Table 1.

It follows from the data in Table 1 that the mechanical work $p_{tr}\Delta V$ during the phase transition of a real rare gas near the triple point is small compared to the change ΔH_{fus} of the internal energy of the constituent atoms. This simplifies the analysis and allows us to neglect the expansion of the system of bound atoms during the phase transition, and hence, to treat this process as a function of only one variable. Below, we consider the temperature range of the melting point where the criterion

$$\Delta H_{fus} \gg p\Delta V \qquad (3.2)$$

is satisfied. Here, p is the pressure on the melting curve and ΔV is the increase of the specific volume. In this temperature range, the attractive part of the pair interaction potential of atoms is responsible for the behavior of this system of atoms, and it is not necessary to explicitly account for the external pressure because that would be very small compared to the attraction forces between the interacting atoms. This criterion is valid under the condition

$$p \ll \frac{D}{R_e^3}. \qquad (3.3)$$

Equation (3.1) and the data in Table 1 imply that these criteria are valid at least at temperatures

$$T < T_{cr}, \qquad (3.4)$$

where T_{cr} is the critical temperature for the liquid–gas transition (see Table 1). We note that at high external pressures, the repulsive part of the pair interaction potential of atoms is important for the phase transition and the mechanical work $p\Delta V$ makes a considerable contribution to the fusion energy.

We now analyze the excitation of a system of many bound atoms with a pair interaction in the case where the interaction between nearest neighbors dominates. Computer simulations of clusters, which are systems of such bonded atoms and have completed shells, show the complex character of the phase transition. At a certain degree of excitation, atoms of filled shells move out of those shells to the cluster surface; they float on it, albeit with somewhat hindered motion, and then return, typically exchanging roles with other atoms coming out of the surface layer [21, 22]. These transitions are easiest for the outermost shell, but are possible for others beneath, and in some range of temperatures one can thus construct several caloric curves [23], each describing the excitation of a particular shell. Evidently, when a cluster becomes very large, one can extract the surface excitation and the bulk (internal) excitation in this way. Although we here draw from the experience of phase transitions in clusters, in order to simplify this analysis, we restrict the further discussion to internal excitations only, and therefore consider infinite clusters or bulk systems of bound atoms.

We note two types of excitations in a bulk system of bound atoms. The first deals with the excitation of vibrations or phonons; this excitation is identical in principle for the solid and liquid states, apart from the regular character of excitations of a periodic lattice. The excitation of the other type, the configuration excitation, corresponds to a change in the atomic positions. When this excitation is small, it can be characterized by the change in the number of vacancies inside the crystal lattice. These vacancies result from removal of atoms from sites of the crystal lattice to positions outside. When the number of vacancies becomes large, such that neighboring vacancies border, these vacancies transform into voids [24], and the energy of formation

of an individual void, as well as its volume, depends on the degree of the configuration excitation. We assume the excitations of these two types to be independent and analyze the configuration excitation, which is responsible for the phase transition as a result of formation of voids inside the system.

Considering the configuration excitation to be unequilibrium with respect to the thermal motion of atoms, we use a simple model for this excitation [25–27]. We prepare an excited state as follows. In a crystal consisting of $n + \nu$ atoms, we create ν vacancies inside by removing ν atoms to the outside. Then ν, the number of vacancies formed, characterizes the excitation degree of this system. In the second stage of the evolution, this system relaxes such that it shrinks, and its internal energy typically (but not necessarily) drops. We characterize the degree of the configuration excitation of this system by the number of voids ν, which coincides with the number of interior vacant sites. Of course, in contrast to a vacancy, an individual void varies its form and volume in time; we consider an individual void in terms of its average form and volume. We use statistical parameters for each void, characterizing it by a certain energy ε_ν of its formation and the statistical weight g_ν. Thus, we describe the degree of the configuration excitation of the system by the number of voids that are initially isolated vacancies. In this manner, we consider an excitation as a gas of noninteracting voids that are identical on average, whereas the parameters of an individual void depend on the degree of excitation.

The statistical model under consideration involves averaging over atomic positions in a system of interacting atoms. If we restrict the treatment to interactions between nearest neighbors only, we can express the excitation energy through the average number of the nearest neighbors n_c for any internal atom. The mean binding energy per atom is then given by

$$\varepsilon = Dn_c/2,$$

where D is the binding energy per individual bond and the average volume per atom is

$$V = 12V_0/n_c,$$

where the specific volume V_0 is that based on the solid state. From this, we have

$$V = V_0\frac{\varepsilon_{\text{sub}}}{\varepsilon} = V_0\frac{n\varepsilon_{\text{sub}}}{\varepsilon_{\text{sub}} - \varepsilon_\nu \nu/n}, \tag{3.5}$$

where ε_ν is the energy of formation of one elementary void and ε_{sub} is the binding energy per atom for the solid state. In particular, this implies that for the liquid state,

$$\xi_{\text{liq}} \equiv \frac{V_0}{V}\frac{\varepsilon_{\text{sub}}}{\varepsilon_{\text{sub}} - \Delta H_{\text{fus}}} = 1, \tag{3.6}$$

and the statistical average of this parameter over real rare gases gives

$$\xi_{\text{liq}} = 1.024 \pm 0.006,$$

and Eq. (3.2) is therefore valid within the accuracy of 3%. Evidently, the error in ξ_{liq} is related to the accuracy of using the mean-field approximation.

Considering the liquid–solid phase transition as a result of formation of voids, we construct the partition function of a void gas as

$$Z(\nu) = C^\nu_{n+\nu} g_\nu^\nu \exp\left(-\frac{\nu\varepsilon_\nu}{T}\right), \tag{3.7}$$

where ε_ν is the energy of formation of an individual void, g_ν is the statistical weight of a void, and these parameters depend on the parameter $x = \nu/n$. Assuming voids to be independent, we define the energy of formation of an individual void as

$$\varepsilon_\nu = \varepsilon_0 - U\left(\frac{\nu}{n}\right), \tag{3.8}$$

where U is the effective interaction potential of voids, ε_0 is the energy of formation of a vacancy in the crystal as a result of removing an internal atom to the surface, i.e., when $\nu = 0$ ($\varepsilon_0 = 6D$ in the case of the interaction between nearest neighbors only). We take the statistical weight of an individual void to have the form [26, 27]

$$g = 1 + a\frac{\nu}{n}, \quad a \gg 1. \tag{3.9}$$

For the effective interaction potential, we use [26, 27]

$$U\left(\frac{\nu}{n}\right) = \varepsilon_0\left[\exp\left(-\frac{\alpha n}{\nu}\right) - \exp\left(-k\frac{\alpha n}{\nu}\right)\right] = \varepsilon_0 u\left(\frac{\nu}{n}\right),$$
$$u(x) = \exp\left(-\frac{\alpha}{x}\right) - \exp\left(-k\frac{\alpha}{x}\right), \tag{3.10}$$

where α and k are parameters. These relations imply that as the unoccupied space inside the system increases, the statistical weight per atom increases, and the energy of formation of new vacancies decreases. It then follows that the reduced logarithm of the partition function is given by

$$\frac{\ln Z(\nu)}{n} = x\ln(1 + ax) + \ln(1 + x)$$
$$+ x\ln\left(1 + \frac{1}{x}\right) - x\frac{\varepsilon_0}{T}[1 - u(x)], \quad x = \frac{\nu}{n}. \tag{3.11}$$

The use of the phase transition parameters for condensed rare gases at their triple points together with this expression for the partition function allows us to find the parameters of this model [25–27]. We note that the complex form (3.6) of the effective interaction potential of voids is related to a bimodal form of the partition function that cannot be realized at $k = \infty$. The parame-

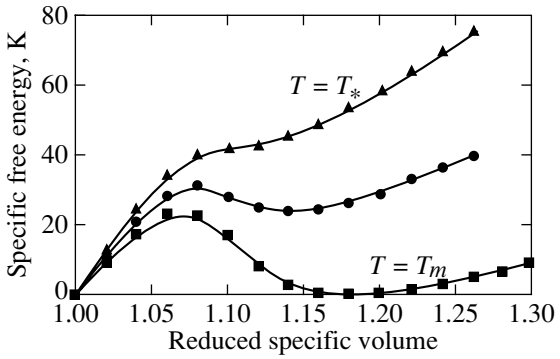

Fig. 3. The dependence of the specific free energy of condensed rare gases on the reduced volume per atom. The right minimum corresponds to the liquid state; at $T = T_*$, the liquid state becomes unstable.

ters of this model averaged over the stable rare gases are given in Table 2 [26, 27].

One can continue the partition function of a system of bound atoms to the range of low temperatures, where the liquid state of the bulk is a metastable aggregate state. Figure 3 shows the dependence of the specific free energy

$$F = -T\ln Z$$

on the specific volume and is based on Eqs. (3.5) and (3.7)–(3.10); the mechanical work as a result of the phase transition is neglected in accordance with criteria (3.3) and (3.4). We use Eq. (3.7) for the specific partition function and the relations between the volume per one atom V, the specific excitation energy $v\varepsilon_v/n$, and the relative number of voids v/n. It follows that, below the freezing limit $T_* = 0.36D$, the liquid maximum of the partition function disappears and the liquid state becomes unstable.

4. THE CALORIC CURVE OF CONDENSED RARE GASES AND THE RATE OF EQUILIBRATION

Thus, neglecting the surface configuration excitation of large clusters of rare gases, we restrict our description to two aggregate states that correspond to the solid and liquid states of a bulk system. Figure 4 gives the caloric curves for these states. Each caloric curve is the temperature dependence of the specific internal energy of the isothermal system of bound atoms. We take the excitation energy as a sum of the phonon excitation energy and the configuration excitation energy. The phonon contribution to the excitation energy per atom is given by [11]

$$E_{\rm ph} = 3Td\left(\frac{\Theta_D}{T}\right), \tag{4.1}$$

where

$$d(x) = \frac{3}{x^3}\int_0^x \frac{z^3 dz}{\exp(z)-1} \tag{4.2}$$

is the Debye function; the Debye temperature Θ_D does not conform to the scaling law. Figure 4 corresponds to argon, for which we take [20, 28] $\Theta_D \approx 90$ K. The specific energy of the configuration excitation is taken as $v\varepsilon_v/n$ for the liquid state and is zero for the solid state, and therefore, the total specific internal energy of the liquid state is

$$E_{\rm liq} = 3Td\left(\frac{\Theta_D}{T}\right) + \frac{v}{n}\varepsilon_{v_{\rm liq}}, \tag{4.3}$$

and the second term is absent for the solid state. Next, the liquid caloric curve terminates at the freezing limit T_* where the liquid maximum of the free energy disappears (see Fig. 3). In the same manner, the solid caloric curve terminates at high temperatures in this formulation. Because the critical temperature of the solid state is sufficiently high such that the mechanical work of solid–liquid transitions cannot be neglected, it would be incorrect to discuss the high-temperature range of the caloric curve within this framework. Because the scaling law is invalid for the phonon excitation energies, we specifically analyze the parameters of argon in what follows.

We note the principal difference in the construction of caloric curves for clusters and for bulk systems. For clusters, the coexistence of the solid and liquid phases is possible in some temperature range where the probability of the location of a cluster in each aggregate state is non-negligible. Hence, assuming that the time intervals between transitions from one aggregate state to another is long compared to the observation time and that the times required for the transitions are brief on the same scale, we terminate the caloric curve of each state at a temperature at which the probability of observing the cluster in this state becomes small. (More rigorously, we could terminate where the local minimum in the free energy of that form disappears.) For large clusters or a bulk system, this probability of the observation of the unfavored phase is very, very small, even in the vicinity of the melting point, but a typical dwell time in one phase may be long, even infinite, pre-

Table 2. Mean parameters of condensed rare gases at the triple point

Quantity	Average value
α	0.13 ± 0.01
k	4.8 ± 0.2
$v_{\rm liq}/n$	0.31 ± 0.01
a	65 ± 15
$U(v_{\rm liq}/n)/D$	3.3 ± 0.2
$U(v_{\rm min}/n)/D$	1.3 ± 0.1
$-\ln Z((v_{\rm min}/n)/n)$	0.39 ± 0.02

cisely at the melting point, if the cluster is macroscopic. Hence, constructing the caloric curve in Fig. 4 for argon, we suppose the mean dwell time in the liquid state to be sufficiently long. The true low-temperature termination of the liquid caloric curve means that at even lower temperatures, the metastable liquid state does not exist. In addition, Fig. 5 gives the specific volume of the argon liquid state; the argon fusion energy for the solid–liquid phase transition is represented in Fig. 6. In accordance with criteria (3.3)–(3.5), our treatment is restricted to a range of temperatures that are not too high.

For a bulk system, the decay of the liquid state below the melting point is determined by fluctuations due to nonuniformities in the void distribution. This also occurs for small clusters. For very large clusters or bulk systems, we neglect these fluctuations, and metastable liquid states can therefore live very long in the absence of external perturbations. Below the freezing limit, the decay of an unstable liquid state results from the diffusion of voids to the boundary of the system; below, we briefly analyze this process.

Because the diffusion process has an activation character, it slows down with a temperature decrease and stops at sufficiently low temperatures. Because the voids move by diffusion, the rate of this process depends on the geometry of the system. We consider atoms to be bonded with a substratum, and bound atoms to form a film on the substratum surface. (The substratum may be another layer of the same material.) If the film thickness is l, the typical time of a void departure outside the film is of the order l^2/D_{dif}, where D_{dif} is the diffusion coefficient of voids inside the film. Because a displacement of voids is determined by the inverse displacement of atoms, the diffusion coefficient of voids is

$$D_{\mathrm{dif}} \sim \omega_D a^2 \exp\left(-\frac{E_a}{T}\right), \tag{4.4}$$

where $\omega_D = \Theta_D/\hbar$ is the Debye frequency, a is the lattice constant, T is a current temperature, and E_a is the activation energy for the atomic displacement, which depends on the relative number of voids or vacancies inside the system. We assume that the heat transport proceeds more effectively than the process of void diffusion because of the activation character of the last process, i.e., the criterion

$$D_{\mathrm{dif}} \ll \chi \tag{4.5}$$

is satisfied, where χ is the thermal diffusivity coefficient.

The activation energy drops if the relative number of voids decreases. Evidently, the "frozen" temperature

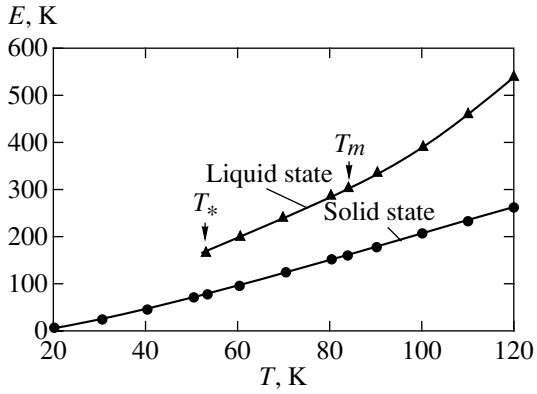

Fig. 4. Caloric curves for argon representing the temperature dependences of the internal energy of aggregate states. The solid caloric curve is given by Eq. (4.1) and the liquid caloric curve is given by Eq. (4.3).

(an analog of the glassy temperature [30]) is determined by the condition

$$\int \frac{D_{\mathrm{dif}}}{l^2} dt \sim 1. \tag{4.6}$$

Assuming the temperature variation rate dT/dt to be constant, we obtain from Eqs. (4.6) and (4.4) that

$$\frac{\omega_D a^2}{l^2} \exp\left(-\frac{E_a}{T_f}\right) \frac{T_f^2}{E_a \dfrac{dT}{dt}} \sim 1, \tag{4.7}$$

where T_f is the temperature below which voids are frozen, and the activation energy $E_a \gg T_f$ corresponds to a frozen relative number of voids.

The activation energy of this process increases as the number of voids decreases. Below, we consider this transition in the limit in which the atoms form a crystal lattice, and even the nearest vacancies do not border each other. We take a face-centered cubic lattice for the solid state of the system of bound atoms. The transition of a vacancy from one lattice site to a neighboring one is similar to the transition of an atom next to a vacancy to the vacancy site. For simplicity, we fix other atoms in the sites of the crystal lattice. To make the transition to a neighboring site, a test atom must overcome a barrier; from symmetry considerations, the barrier height is the difference of the total interaction potentials of atoms with the test atom located in a site of the crystal lattice and halfway between two neighboring vacancies. If we introduce the pair interaction potential of atoms $U(R)$ at the distance R between them, we obtain the barrier height

$$E_a = -11U(a) - 2U(\sqrt{2}a) - 4U(\sqrt{3}a) - U(2a) + 4U\left(\frac{\sqrt{3}}{2}a\right) + 4U\left(\frac{\sqrt{5}}{2}a\right) + 8U\left(\frac{\sqrt{7}}{2}a\right) + 2U\left(\frac{3}{2}a\right). \tag{4.8}$$

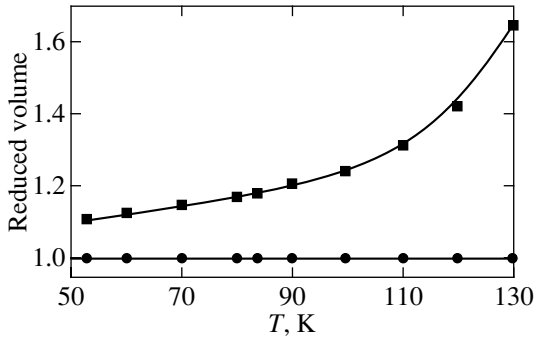

Fig. 5. The volume per atom for solid and liquid argon states in neglecting the anharmonicity of the atom motion.

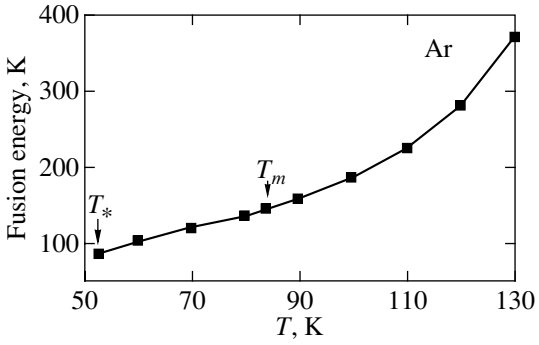

Fig. 6. The fusion energy for the solid–liquid phase transition in argon.

We account for the interaction of a transiting atom with the nearest neighbors as it passes from the initial to the final atom position; a is the distance between nearest neighbors of the lattice.

In particular, we use the Lennard-Jones interaction potential between atoms

$$U(R) = D\left[\left(\frac{R_e}{R}\right)^{12} - 2\left(\frac{R_e}{R}\right)^{6}\right], \tag{4.9}$$

where R_e is the equilibrium distance between atoms for a classical diatomic molecule and D is its dissociation energy. In this case, we have

$$E_a = 9.2D. \tag{4.10}$$

For the Morse interaction potential

$$U(R) = D[\exp(-2\alpha(R-R_e)) - 2\exp(\alpha(R-R_e))] \tag{4.11}$$

and the Morse parameter $\alpha = 6/R_e$ (making the potential as similar to the Lennard-Jones one as possible), we have

$$E_a = 8.2D. \tag{4.12}$$

We note that both interaction potentials are characterized by identical dissociation energies D of the diatomic molecule and identical second derivatives of the interaction potentials at their equilibrium distances,

$$U''(R_e) = 72/R_e^2.$$

If we restrict the interactions to those between nearest neighbors of the crystal lattice, then $a = R_e$. These results can be used to obtain an upper limit for the activation energy of the transition under consideration. The criterion of validity of Eqs. (4.11) and (4.12) is such that neighboring voids are individual vacancies at the crystal lattice sites, i.e., $v \ll n/12$.

We now analyze the character of the frozen process under the conditions of the specific experiment [29, 30], when the thickness of the argon film on the copper target is 0.1 μm and the cooling rate is 2 K/min. In accordance with Eq. (4.7), voids are frozen if $E_a/T_f \approx 14$. If the decay of an unstable state resulting from an irreversible transport of voids starts from the temperature $T_* = 52$ K, some fraction of the voids diffuses to the outside under the given conditions, until the activation energy of the void diffusion process reaches the value

$$E_a = 14T_* \approx 5D.$$

Because this value is less than the activation energy (4.10) and (4.12) for the diffusion of vacancies in the crystal, some of the voids are frozen by this cooling process. Thus, it follows from the above estimate that a system of bound atoms is characterized by a nonequilibrium number of voids or vacancies that are caught and frozen at low temperatures, and this number depends on the rate of cooling of this system.

5. CONCLUSION

Although there is no difference in the forms of the temperature dependence of some physical parameters for systems of a finite number of bound atoms in the cases of chemical transformations and phase transitions, these phenomena are different in principle due to the different dependences of the corresponding partition functions on the number of elementary excitations or the excitation energy. For chemical transformations, the partition function has a sharp maximum at the average number of excitations at a given temperature, and this maximum tends to infinity as the number of excitations tends to infinity. The temperature variation leads to a smooth transition from one chemical state of the system to the other. For the phase transition of a large system, when surface excitations are not important, the partition function has two maxima as a function of the number of elementary configuration excitations; these maxima correspond, e.g., to the solid and the liquid. The liquid state is characterized by a freezing limit below which the liquid maximum disappears. Quenching the configuration excitations in a large system of interacting atoms results in a transport of voids to the boundaries of the system or from them. The character

of the void transport process determines the state of this system of bound atoms after its cooling.

ACKNOWLEDGMENTS

This paper was supported in part by the Russian Foundation for Basic Research (project no. 99-02-16094) and by the National Science Foundation (USA).

REFERENCES

1. R. Brout, *Phase Transitions* (Benjamin, New York, 1965).
2. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena* (Oxford Univ. Press, New York, 1971).
3. A. R. Ubbelohde, *The Molten State of Matter* (Wiley, Chichester, 1978).
4. R. S. Berry, J. Jellinek, and G. Natanson, Phys. Rev. A **30**, 919 (1984).
5. R. S. Berry, J. Jellinek, and G. Natanson, Chem. Phys. Lett. **107**, 277 (1984).
6. J. Jellinek, T. L. Beck, and R. S. Berry, J. Chem. Phys. **84**, 2783 (1986).
7. R. S. Berry, T. L. Beck, H. L. Davis, and J. Jellinek, Adv. Chem. Phys. **90**, 75 (1988).
8. D. J. Wales and R. S. Berry, J. Chem. Phys. **92**, 4283 (1990).
9. R. S. Berry, Nature **393**, 238 (1998).
10. R. S. Berry, in *Theory of Atomic and Molecular Clusters*, Ed. by J. Jellinek (Springer-Verlag, Berlin, 1999), p. 1.
11. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, *Statistical Physics* (Nauka, Moscow, 1995; Pergamon, Oxford, 1980), Vol. 1.
12. B. M. Smirnov, *Physics of Ionized Gases* (Wiley, New York, 2001).
13. R. A. Aziz and M. J. Slaman, Chem. Phys. **130**, 187 (1989).
14. R. A. Aziz and M. J. Slaman, J. Chem. Phys. **92**, 1030 (1990).
15. A. K. Dham, A. R. Allnatt, W. J. Meath, and R. A. Aziz, Mol. Phys. **67**, 1291 (1989).
16. A. K. Dham, W. J. Meath, A. R. Allnatt, *et al.***,** Chem. Phys. **142**, 173 (1990).
17. B. M. Smirnov, Usp. Fiz. Nauk **162** (12), 97 (1992) [Sov. Phys. Usp. **35**, 1052 (1992)].
18. B. M. Smirnov, Usp. Fiz. Nauk **164**, 1165 (1994) [Phys. Usp. **37**, 1079 (1994)].
19. B. M. Smirnov, *Clusters and Small Particles in Gases and Plasmas* (Springer-Verlag, New York, 2000).
20. Ch. Kittel, *Introduction to Solid State Physics* (Wiley, New York, 1986), 6th ed.
21. H. P. Cheng and R. S. Berry, Phys. Rev. A **45**, 7969 (1992).
22. H. P. Cheng, X. Li, R. L. Whetten, and R. S. Berry, Phys. Rev. A **46**, 791 (1992).
23. R. E. Kunz and R. S. Berry, Phys. Rev. E **49**, 1895 (1994).
24. H. Reiss, H. L. Frisch, and J. L. Lebowitz, J. Chem. Phys. **31**, 369 (1959).
25. B. M. Smirnov, Phys. Scr. **58**, 595 (1998).
26. B. M. Smirnov, Inorg. Mater. **35**, 562 (1999).
27. B. M. Smirnov, in *Nucleation Theory and Applications*, Ed. by J. W. R. Schmelzer, G. Röpke, and V. B. Priezzhev (Joint Inst. for Nuclear Research, Dubna, 1999), p. 355.
28. N. Schwenther, E. E. Koch, and J. Jortner, *Electronic Excitations in Condensed Rare Gases* (Springer-Verlag, Berlin, 1985).
29. A. Kouchi and T. Kuroda, Jpn. J. Appl. Phys. **29**, 807 (1990).
30. I. Gutzow and J. Schmelzer, *The Vitreous State* (Springer-Verlag, Berlin, 1995).

Journal of Experimental and Theoretical Physics, Vol. 93, No. 4, 2001, pp. 786–795.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 899–909.

SOLIDS
Structure

Short-Range Magnetic Order in a Spin Density Wave in Cr-Based Multilayers

V. N. Men'shov* and V. V. Tugushev**

Russian Research Centre Kurchatov Institute, pr. Kurchatova 1, Moscow, 123182 Russia
**e-mail: sasha@mics.msu.su*
***e-mail: vvtugushev@mail.ru*
Received April 12, 2001

Abstract—A mechanism of the formation of the short antiferromagnetic order with a spin density wave (SDW) in the vicinity of the interfaces in the Fe/Cr type multilayers is proposed. The main reason behind the emergence of magnetic ordering with SDWs is the redistribution of charge (and, hence, spin) density in the vicinity of Fe/Cr interfaces, which leads to the paramagnetic phase instability at a temperature considerably higher than the Néel temperature in chromium. The Ginzburg–Landau expansion for the free energy of the system is used for determining the inhomogeneous collinear structures of CDWs and for constructing the phase diagram (the dependence of the transition temperature on the thickness of the antiferromagnetic interlayer). The obtained results are used for discussing the experimental data on neutron scattering and tunnel microscopy. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The peculiar magnetic and kinetic properties of a number of artificial layered structures on the basis of transition metals (multilayers having a thickness from tens of angstroms to several micrometers) have attracted considerable attention and have become an object of heated discussions in recent years. We are speaking, in particular, of the NM/AFM and the FM/AFM structures, where NM, FM, and AFM stand for the normal, ferro- and antiferromagnetic metals, respectively, and chromium and its diluted alloys are used predominantly as AFM. In view of certain technological and practical factors associated primarily with the application of such structures in magnetic recording systems, Fe/Cr multilayers whose properties are considered in recent reviews [1, 2], have become most popular. Among other things, these multilayers were first examples of systems with a sign-oscillating effective exchange between the FM (Fe) layers and the AFM (Cr) interlayer depending on the thickness of the latter [3].The complex dependence of this exchange on the interlayer thickness and the quality of the Fe/Cr interface (the presence of "short" and "long" periods), the giant magnetoresistance, the peculiar magnetic phase diagram obtained from neutron-diffraction and magnetooptical measurements, the emergence of collinear as well as complex noncollinear structures, and a number of other experimental results have not received an adequate theoretical interpretation. It can only be stated undeniably that the band nature of the magnetism in the AFM layer and the special sensitivity of this type of ordering to the doping level, temperature, and external effects is of fundamental importance for the properties of Fe/Cr type multilayers.

It should be recalled that, according to the generally accepted point of view, a peculiar magnetic state of band electrons, viz., spin density wave (SDW) [4], is realized in bulk chromium below the Néel temperature T_N = 311 K. This state is characterized by transverse-polarized spin density distribution (which becomes longitudinally polarized below the spin-flip transition temperature T_{SF} = 123 K),

$$\mathbf{S}(\mathbf{r}) = \boldsymbol{\sigma}(x)\cos(\mathbf{Q}\cdot\mathbf{r}), \qquad (1)$$

with the wave vector $\mathbf{Q} = (2\pi/a)(100)$, where a is the bcc lattice constant. The SDW amplitude $\sigma(x) = \mathbf{n}\sigma(x)$, where $\mathbf{n} \perp \mathbf{Q}$ in a transverse-polarized wave and $\mathbf{n} \parallel \mathbf{Q}$ in a longitudinally polarized wave ($\mathbf{n}$ is a unit vector); $\sigma(x)$ is a complex long-period function whose period λ is incommensurate with a. The specific form of the function $\sigma(x)$ is either preset (in simple empirical or variational algorithms, it is simply assumed that $\sigma(x) = \sigma_0\cos(qx)$, $q = 2\pi/\lambda$, $\lambda \gg a$ [5]) or calculated self-consistently by minimizing the free energy (for example, in the model of the soliton lattice of the SDW, the dependence $\sigma(x)$ has the form of an elliptic function [6]).

The magnetic phase diagrams of bulk chromium and its numerous diluted alloys have been investigated comprehensively (see, for example, review [7]). The magnetic phase diagram of the Fe/Cr layered structure has also been established in general [8–10]. In the latter case, we are speaking of the dependence of the temperature of the antiferromagnetic transition in the Cr interlayer on its thickness L. A qualitative distinguishing feature of all available "bulk" phase diagrams is that

two temperatures, $T_1(L)$ and $T_N(L)$, of transition to an antiferromagnetic state of the SDW type are observed for a fixed composition of the interlayer with a thickness $L > L^* \approx 30$ monolayers (approximately 45 Å). For $L < L^*$, the temperature $T_N(L)$ is equal to zero, and the temperature $T_1(L)$ remains finite (approximately equal to 600 K), while for $L \gg L^*$, the temperature $T_N(L)$ tends to the "bulk" value $T_N(\infty) = 311$ K and $T_1(L)$ varies slowly near $T_1(\infty) \approx 550$ K.

The nature of the magnetic transition to the "high-temperature" phase and the existence of the "critical" thickness remain the subjects of numerous discussions, while the transition to the low-temperature phase is reliably associated with ordering of the type of the transition to the state with an incommensurate SDW (ISDW) in bulk chromium. In accordance with the standard theory of one-parametric scaling, the dependence $T_N(L)$ should have the form

$$\frac{T_N(L) - T_N(\infty)}{T_N(L)} = bL^{-1/\nu},$$

where b is a constant and ν is the critical exponent for correlation length, which naturally depends on the type of magnetic order [11]. However, the experiments give another dependence $T_N(L)$ resembling that given above, but with the effective length $L_{\text{eff}} = L - L^*$ for $L > L^*$ ($T_N(L) = 0$ for $L < L^*$). It is more or less clear that the thickness L^* and the transition at $T_1(L)$ are associated with the effects occurring in the vicinity of the Fe/Cr interface, although this problem has not been analyzed in detail so far. For some not quite clear reasons, the transition at $T_1(L)$ is associated in the literature with the emergence of a commensurate SDW (CSDW); however, in the subsequent analysis, we will try to avoid the terms "commensurate and incommensurate SDW" which are applicable to bulk systems and leave the popular identification of $T_1(L)$ with the temperature of transition to a phase with a commensurate SDW without comments.

It was proposed in [12] that below the temperature $T_1(L)$, multilayers of the Fe/Cr type experience an antiferromagnetic ordering to an inhomogeneous phase of SDW which has no bulk analogue and cannot be described using the generally accepted terms of commensurate or incommensurate SDW structures. The origin of this new phase is determined completely by the existence of the Fe/Cr interfaces and by the redistribution of the charge and spin densities of band electrons in the vicinity of these interfaces. Such a "high-temperature" SDW phase can be interpreted qualitatively as a short- range antiferromagnetic order in the Cr interlayer, which is "induced" by the interfaces and characterized by a new scale, viz., the interpolation length L^*.

It should be noted that the naive attempt made in [13] to apply the simplest variational non-self-consistent approach entirely based on the analogy with the algorithms used for bulk systems with SDWs did not make it possible to explain the magnetic phase diagram of Fe/Cr multilayers even qualitatively, not to mention the fine details of the spin density distribution in the Cr interlayer and the competition between collinear and noncollinear structures with SDWs. A numerical analysis of the magnetization distributions in the layers [14] makes it possible, in principle, to describe the ground state of the Fe/Cr structure for various thicknesses L, but it cannot be used for studying the thermodynamics and complex magnetic configurations in regions with $T < T_1(L)$ (for "thin" structures with $L < L^*$) and $T_N(L) < T < T_1(L)$ (for "thick" structures with $L > L^*$). Unfortunately, we cannot claim to providing a quantitative analysis of the (T, L) phase diagrams for multilayers of the Fe/Cr type in the entire range of temperatures T and thicknesses L; however, even the qualitative approach which will be described below for the short-range order region is of considerable interest in our opinion in view of the peculiarity and variety of the antiferromagnetic structures formed in this region.

2. FORMULATION OF THE PROBLEM

Let us consider a simplified 1D model of the typical element of a Fe/Cr multilayered structure, i.e., a three-layered system consisting of two outer layers of the ferromagnetic metal (Fe) and an interlayer of the weakly antiferromagnetic metal (Cr), oriented along the **x** direction perpendicular to the interfaces. We assume that Fe/Cr surfaces are perfectly smooth and arranged symmetrically (at distances $\pm l$) relative to the origin $x = 0$ at the middle of the interlayer. The antiferromagnetic order parameter $\mathbf{\Delta}(x)$ is introduced in the standard manner (see, for example, review [15]) and is assumed to be transverse-polarized relative to the x axis of the structure:

$$\mathbf{\Delta}(x) = U\boldsymbol{\sigma}(x),$$

where $\boldsymbol{\sigma}(x)$ is the SDW amplitude (1) mentioned in the Introduction and U is the effective SDW potential which generally depends on the choice of the microscopic model (see [5]). We present the order parameter in the form

$$\mathbf{\Delta}(x) = \mathbf{n}_y\Delta_y(x) + \mathbf{n}_z\Delta_z(x), \quad \mathbf{\Delta}(x) \perp \mathbf{n}_x, \qquad (2)$$

where $\mathbf{n}_x$, $\mathbf{n}_y$, $\mathbf{n}_z$ are the unit vectors of polarization and $|x| < l$, l being the half-thickness of the interlayer. Since the Curie temperature in the Fe layers ($T_C \approx 1040$ K) is much higher than the characteristic temperatures of SDW formation in the Cr interlayer, we assume that the magnetic moment in the Fe layers is homogeneous and independent of temperature.

In this work, we consider the temperature region above the Néel temperature in the bulk of the antiferromagnet ($T > T_N(\infty)$); i.e., we assume the absence of a long-range magnetic order. At the same time, we assume that both quantum-mechanical and classical

thermodynamic fluctuations of the order parameter $\mathbf{\Delta}(x)$ are small (the size of the system, including that along the x axis, is much larger than the atomic spacing and the effective interaction constant is much smaller than unity). In other words, the conditions are satisfied for analyzing the thermodynamics of the system on the basis of the theory of a "local phase transition" [15, 16] using formally the Ginzburg–Landau free energy functional and applying the mean-field approximation for its analysis. The estimation of the correlation length of the order parameter for the standard SDW model in chrome and its alloys [15] gives the values

$$\xi(T) = \frac{\xi_0}{\sqrt{T/T_N(\infty) - 1}},$$

where $\xi_0 \approx (5–10)a$, $a \approx 2.8$ Å is the parameter of bcc lattice of chromium. The characteristic temperatures considered here lie in the interval $T_N(\infty) < T < T_1(L)$, $T_1(L) \approx 600$ K, $T_N(\infty) \approx 311$ K; the characteristic thicknesses $L = 2l$, where $l \geq \xi(T) \geq \xi_0 \gg a$. Thus, the mean-field approach considered by us can be regarded as satisfactory for not very thin interlayers and for relatively high temperatures (which are still lower than the Curie temperature in the ferromagnetic layers). The behavior of the order parameter in the vicinity of Fe/Cr interfaces on scales smaller than ξ_0 is not described in this approach and is simply defined through the boundary conditions under certain reasonable assumptions.

Let us formulate in greater detail the proposed approach on the basis of the Ginzburg–Landau expansion of the free energy F of the system. We assume that in the temperature range under investigation, the order parameter $\mathbf{\Delta}(x)$ is small ($|\mathbf{\Delta}(x)| \ll \pi T$) and varies slowly with coordinate x ($|\partial\mathbf{\Delta}/\partial x| \ll \pi T/\xi_0$) so that the series expansion of the functional $F[\mathbf{\Delta}(x)]$ in $\mathbf{\Delta}(x)$ and in its derivatives is valid. For thick interlayers ($L > 2\xi_0$), we disregard the details of the spatial variation of $\mathbf{\Delta}(x)$ over distances smaller than ξ_0 near the Fe/Cr interfaces, where the local approximation for the functional $F[\mathbf{\Delta}(x)]$ is inapplicable, and write the latter as the sum of the "volume" and "surface" components:

$$F = F_V + F_S. \tag{3}$$

Here, the volume contribution is defined as

$$F_V = \frac{1}{2}\int_{-l}^{l} f_V[\mathbf{\Delta}(x)]dx, \tag{4}$$

$$f_V[\mathbf{\Delta}(x)] = c_1\mathbf{\Delta}^2 + c_2 v_F^2\mathbf{\Delta}'^2 + c_2\mathbf{\Delta}^4, \tag{5}$$

and the surface contribution is given by

$$F_S = \frac{\nu}{4}[\mathbf{\Delta}^2(l) + \mathbf{\Delta}^2(-l)] + \frac{1}{2}[\mathbf{A}(l)\cdot\mathbf{\Delta}(l) + \mathbf{A}(-l)\cdot\mathbf{\Delta}(-l)], \tag{6}$$

where v_F is the Fermi velocity and ν and $\mathbf{A}$ are parameters. Strictly speaking, the decomposition of F into the volume and surface components in the form (3)–(6) is incorrect for systems with thin interlayers ($L < 2\xi_0$).

Expressions (3)–(6) can be derived directly from the microscopic model of SDW by the standard method of Green's functions using the high-temperature expansion of the diagrammatic series for free energy (see, for example, review [15] and the expression for $f_V[\mathbf{\Delta}(x)]$ given therein). It should be noted that in the temperature range $T \approx T_N(\infty)$, we should use, instead of the simplified expansion (5) for $f_V[\mathbf{\Delta}(x)]$, a more complex expression containing higher-order terms proportional to $\mathbf{\Delta}^6(x)$ (and $\mathbf{\Delta}''^2(x)$). This is associated with specific features of the microscopic model with nesting, in which the coefficients of the terms proportional to $\mathbf{\Delta}^4(x)$ and $\mathbf{\Delta}'^2(x)$ in formula (5) changes their sign in the temperature range $T \sim T_N$; this leads to complex self- consistent equations for $\mathbf{\Delta}(x)$ since higher-order terms in $\mathbf{\Delta}(x)$ and $\mathbf{\Delta}'(x)$ in functional F_V must be taken into account in this case [15]. In the region of temperatures much higher than $T_N(\infty)$ we are interested in, all coefficients in relation (5) are positive and there is no need to write higher-order terms.

The physical meaning of the first and second contributions to the surface energy F_S is quite clear and is of the general nature. Namely, the term linear in $\mathbf{\Delta}$ is directly connected with the effective exchange interaction between electron spins of the outer ferromagnetic layers (Fe) and the antiferromagnetic interlayer (Cr) in the vicinity of interfaces. The term quadratic in $\mathbf{\Delta}$ is due to the conventional redistribution of charge density and the presence of the contact potential difference near the surface separating the Fe and Cr layers. As a result, the electron polarizability, the electron spectrum parameters, and the band filling in the vicinity of the Fe/Cr interfaces change; the conditions of instability of the paramagnetic phase relative to the SDW formation in the surface layer of chromium change accordingly. The estimates of parameters ν and $\mathbf{A}$ in expression (6) can be obtained from the formulas given in [16] for the model with the nesting of the electron and hole regions on the Fermi surface. If U_0 and J_0 are the Coulomb and the exchange components of the surface potential, respectively ($U_0 < 0$ and $J_0 > 0$ in the chosen model, which corresponds to the flow of electrons from Fe to Cr in the surface layer and to the antiferromagnetic interaction of spins in Fe and Cr), we obtain the following estimates in the first order in U_0 and J_0 in the local approximation, assuming that $\mathbf{\Delta}(x)$ is a slowly varying function over the radius of action of the surface potential:

$$\nu \propto -U_0\bar{N}^2/n_0, \quad A \propto J_0 S\bar{N}. \tag{7}$$

Here, $\bar{N}$ is the average density of states at the Fermi level for spectral regions with nesting, $n_0 < 0$ is the dif-

ference in the filling of the electron and hole regions, and S is the spin density per unit interface in Fe. The first (Coulomb) term in interaction (6) prevails when

$$|\mathbf{\Delta}| \gg n_0 \bar{N}^{-1}(J_0 S/U_0),$$

while the second (exchange) term dominates for

$$|\mathbf{\Delta}| \ll n_0 \bar{N}^{-1}(J_0 S/U_0).$$

Since the order parameter $\mathbf{\Delta}$ in the Ginzburg–Landau expansion must be small compared to πT, the role of the term quadratic in $\mathbf{\Delta}$ in expression (6) must be analyzed in greater detail. In chromium, $|n_0| \approx 0.05$ and $\bar{N}^{-1} \approx 1$ eV so that $|J_0 S/U_0| \ll 1$ and the Coulomb term in the temperature range $T \approx 400$–600 K may be larger than the exchange term even if $|\mathbf{\Delta}| \ll \pi T$. It was mentioned above that potential U_0 is associated with the charge flow between Fe and Cr and can be estimated in the order of magnitude as

$$U_0 \approx 4\pi Q e/d$$

(Q is the surface charge and d is the Debye screening length in Cr), i.e., at about 1 eV. It is more difficult to estimate the exchange component J_0 in view of the complex nature of the spin density distribution in the vicinity of the Fe/Cr(100) interface. As a matter of fact, the first monolayer of chromium at this interface is actually rearranged to a considerable extent and, apparently, a strongly localized moment (approximately equal to $1.5\mu_B$ per Cr atom) antiparallel to the ferromagnetic moment of Fe is formed in it instead of the band moment [2]. As a result of such a strong rearrangement, the exchange interaction between the spin of Fe and the band component of the spin density in the inner layers of the Cr interlayer forming the SDW are effectively screened. It is well known that the effective screening in diluted $Cr_{1-x}Fe_x$ alloys strongly suppresses the interaction between the local atomic spin of Fe and the SDWs, which is manifested virtually in a complete absence of "freezing" of this spin in the antiferromagnetic phase down to extremely low temperatures [4, 17]. Estimating of the parameter $|J_0 S/U_0|$ in bulk diluted $Cr_{1-x}Fe_x$ alloys gives values of the order of $(0.01$–$0.1) \ll 1$ and we have no reasons for a considerable departure from such values in the case of the Fe/Cr interface. Thus, in the model expression (6) for functional F_S, the role of the term quadratic in $\mathbf{\Delta}$ is quite significant in our opinion and is probably even more important that the role of the term linear in $\mathbf{\Delta}$. After all these essential explanations, we directly go over to an analysis of the SDW structure and the phase diagram for the model under investigation.

3. STRUCTURE OF THE LONG-RANGE ANTIFERROMAGNETIC ORDER INDUCED BY THE CHARGE DENSITY REDISTRIBUTION IN THE VICINITY OF THE INTERFACE

In our opinion, the arguments given in Section 2 are serious enough to consider the situation with the exchange contribution to F_S, which is linear in $\mathbf{\Delta}(\pm l)$, formally equal to zero as the initial approximation in the problem of the formation of the short-range order in SDW, bearing in mind that this approximation is violated for $|\mathbf{\Delta}| \ll (A/|\nu|) \ll \pi T$. Thus, retaining only the contribution to F_S which is quadratic in $\mathbf{\Delta}(\pm l)$ and is associated with the charge density redistribution near the Fe/Cr interfaces, we carry out the variation of functional (3) in $\mathbf{\Delta}(x)$ in order to find the optimal (in energy) configuration of the SDW in the mean-field approximation. It can easily be verified that the self-consistent equation has the conventional form

$$v_F^2 c_2 \mathbf{\Delta}'' - c_1 \mathbf{\Delta} - 2c_2 \mathbf{\Delta}^3 = 0 \qquad (8)$$

with the boundary conditions

$$\mathbf{\Delta}'(\pm l) \mp \frac{|\nu|}{2c_2 v_F^2}\mathbf{\Delta}(\pm l) = 0. \qquad (9)$$

In this work, we will confine our analysis only to the case of a linearly polarized SDW, for which $\mathbf{\Delta}(x) = \mathbf{n}_z \Delta(x)$ in formula (2), i.e., $\Delta_y(x) = 0$, and $\Delta_z(x) = \Delta(x)$ is the scalar function. In this case, the solutions of the system of equations (8) and (9) can be obtained in the class of the Jacobi elliptic functions since the first integral of Eq. (8) is well known. In the standard notation [18], we will write the possible solutions (one symmetric and two antisymmetric) of the problem (8), (9), which will be subsequently analyzed. Thus, the symmetric solution $\Delta_+(x)$ has the form

$$\Delta_+(x) = \frac{v_F}{\xi}\frac{k'}{\sqrt{2k^2-1}}\mathrm{nc}\left(\frac{1}{\sqrt{2k^2-1}}\frac{x}{\xi}, k\right), \qquad (10)$$

where $\xi = v_F\sqrt{c_2/c_1}$ is the correlation length and k is the modulus of the elliptic function, which depends on the characteristic lengths of the system and temperature:

$$\mathrm{dc}\left(\frac{1}{\sqrt{2k^2-1}}\frac{1}{\xi}, k\right)\mathrm{sn}\left(\frac{1}{2k^2-1}\frac{1}{\xi}\right) = \sqrt{2k^2-1}\frac{\xi}{D}. \qquad (11)$$

Here, $D = 2c_2 v_F^2/|\nu|$ is the new characteristic length (the so-called interpolation length) emerging due to the presence of the term associated with the charge density redistribution and proportional to $\Delta^2(\pm l)$ in formula (5) for surface energy. It should be noted that this length in

the microscopic model used here is almost independent of temperature:

$$D \approx \frac{v_F^2 \bar{N}}{d n_0 U_0} \sim \frac{\varepsilon_F a}{n_0 U_0} \gg a,$$

if $|n_0| \ll 1$ and $|\varepsilon_F/U_0| \geq 1$, which is presumed by the condition of the problem. The relation between ξ and D may in principle be arbitrary and depends on temperature, and parameter k obviously lies in the interval $1 > k^2 > 1/2$, while $l/\xi < \sqrt{2k^2 - 1}\,K(k)$, where $K(k)$ is the total elliptical integral of the first kind and $k' = \sqrt{1 - k^2}$ is an additional modulus.

In addition to the symmetric solution (10) and (11), there exist two antisymmetric solutions $\Delta_-^{(1,2)}(x)$. The first of them has the form

$$\Delta_-^{(1)} = \frac{v_F}{\xi}\frac{k'}{\sqrt{2-k^2}}\mathrm{sc}\left(\frac{1}{\sqrt{2-k^2}}\frac{x}{\xi}, k\right), \tag{12}$$

where parameter k is defined by the equality

$$\mathrm{dc}\left(\frac{1}{\sqrt{2-k^2}}\frac{l}{\xi}, k\right)\mathrm{ns}\left(\frac{1}{\sqrt{2-k^2}}\frac{l}{\xi}, k\right) = \sqrt{2-k^2}\frac{\xi}{D}, \tag{13}$$

and $0 < k^2 < 1$, $l/\xi < \sqrt{2-k^2}\,K(k)$. The second solution has the form

$$\Delta_-^{(2)} = \frac{v_F}{\xi}\frac{1}{\sqrt{2(1-2k^2)}}\mathrm{sc}\left(\frac{1}{\sqrt{2(1-2k^2)}}\frac{x}{\xi}, k\right) \times \mathrm{dn}\left(\frac{1}{\sqrt{2(1-2k^2)}}\frac{x}{\xi}, k\right), \tag{14}$$

where parameter k can be determined from the condition

$$\mathrm{cd}\left(\frac{1}{\sqrt{2(1-2k^2)}}\frac{l}{\xi}, k\right)\mathrm{ns}\left(\frac{1}{\sqrt{2(1-2k^2)}}\frac{l}{\xi}, k\right) \times \left[k'^2\mathrm{nc}^2\left(\frac{1}{\sqrt{2(1-2k^2)}}\frac{l}{\xi}, k\right) + k'^2\mathrm{cn}^2\left(\frac{1}{\sqrt{2(1-2k^2)}}\frac{l}{\xi}, k\right)\right] = \sqrt{2(1-2k^2)}\frac{\xi}{D}, \tag{15}$$

and $0 < k^2 < 1/2$, $l/\xi < \sqrt{2(1-2k^2)}\,K(k)$.

The limiting transition $k \longrightarrow 1$ ($k' \longrightarrow 0$) in formulas (11) and (13) defines the stability boundary for the paramagnetic state for the emergence of SDW of the appropriate symmetry. In the limit $k' \longrightarrow 0$, we have

$$\Delta_+(x) = \frac{v_F}{\xi}k_s'\cosh\frac{x}{\xi}, \tag{16}$$

$$\Delta_-^{(1)}(x) = \frac{v_F}{\xi}k_a'\sinh\frac{x}{\xi}, \tag{17}$$

where parameters $k_{s,a}'$ are functions of l, D, and ξ:

$$k_{s,a}'^2 = \left[\frac{\xi}{D}\left(\coth\frac{l}{\xi}\right)^{\pm 1} - 1\right]/f_\pm\left(\frac{l}{\xi}\right),$$
$$f_\pm(y) = \pm 1 + \frac{1}{4}\cosh(2y) + \frac{3y}{2\sinh(2y)}, \tag{18}$$

the sign + (–) corresponding to the solution of the symmetric (antisymmetric) type. It should be noted that solution $\Delta_-^{(2)}(x)$ in the limit $k' \longrightarrow 0$ under consideration is absent. For $k_{s,a}'^2 = 0$, we obtain from relation (18) the equations for the temperatures $T_{s,a}(l)$ of the transition to the corresponding states with SDWs. In particular, for the symmetric state, this means that

$$\xi^2 = Dl, \quad l \ll \xi,$$
$$\xi = D\left[1 - 2\exp\left(-\frac{2l}{D}\right)\right], \quad l \gg \xi.$$

It is convenient to introduce the dimensionless transition temperature $\tau_s(l) = T_s(l)/T_N^0 > 1$, where T_N^0 is the transition temperature in the model with ideal nesting. This leads to the following estimates for the dependence $\tau_s(l)$ in the limiting cases:

$$\tau_s^2(l)\ln[\tau_s(l)] \approx \begin{cases} \dfrac{D}{l}\tau_0^2, & l \ll D \\ \tau_0^2\left[1 + 4\exp\left(-\dfrac{2l}{D}\right)\right], & l \gg D, \end{cases} \tag{19}$$

where $\tau_0 = v_F/\pi D T_N^0$. Thus, the transition temperature $\tau_s(l)$ for a thick interlayer ($l \gg D$) tends to a finite value $\tau_s(\infty) > 1$ which is independent of the thickness. For $\tau_0 \gg 1$, this value can be estimated with the logarithmic accuracy as

$$\tau_s(\infty) \sim \tau_0/\sqrt{\ln\tau_0},$$

while for $\tau_0 \ll 1$, the value of $\tau_s(\infty) - 1 \sim \tau_0^2$. For a thin interlayer ($l \ll$ D), we can write with the logarithmic accuracy

$$\tau_s^2(l) \sim \frac{\tau_0^2 D}{l\ln(D/l\tau_0^2)},$$

naturally, within the range of application of the Ginzburg–Landau expansion in the form (4)–(6) and the validity of the mean-field approximation (see above).

For an antisymmetric state, a nontrivial solution of problem (8) and (9) exists only for $l > D$ (it should be

recalled that we consider only the region above the bulk Néel temperature so that the value of $\xi(T)$ is finite in the temperature range under investigation). Introducing the dimensionless transition temperature $\tau_s(l) = T_a(l)/T_N^0$, we can estimate, using relation (18) for $k_a'^2 = 0$, the dependence $\tau_a(l)$ in the limiting cases:

$$\tau_a(l) - 1 \approx 3\tau_0^2\left(\frac{l}{D} - 1\right), \quad 0 < \frac{l}{D} - 1 \ll 1,$$
$$\tau_a^2(l)\ln[\tau_a l)] \approx \tau_0^2\left[1 - 4\exp\left(-\frac{2l}{D}\right)\right], \quad \frac{l}{D} \gg 1, \tag{20}$$

and formally assume that $\tau_a(l) = 1$ for $l/D < 1$. Obviously, the temperature $\tau_s(l) > \tau_a(l)$ in the entire range of the ratios l/D, the values of $\tau_a(l)$ and $\tau_s(l)$ approaching each other exponentially for $l \gg D$. Thus, the symmetric solution (10) always appears at a higher temperature than the antisymmetric solution (12) of the first type.

Let us now consider the antisymmetric solution (14) of the second type. It can easily be seen that it is transformed into the solution (12) of the first type for $k = 0$; in accordance with relations (13) and (15), this gives the transition line between these antisymmetric solutions:

$$\frac{\sqrt{2}D}{\xi} = \sin\left(\frac{\sqrt{2}l}{\xi}\right), \quad 0 < \frac{\sqrt{2}l}{\xi} < \pi. \tag{21}$$

The solution of Eq. (21) exists only for $l > D$, and $\xi \longrightarrow \infty$ both for $l/D \gg 1$ and for $l/D \longrightarrow 1$. The corresponding transition temperature $T_{aa}(l)$ defined by relation (21) always lies in the interval $T_N^0 < T_{aa}(l) < T_a(l)$. Omitting a detailed cumbersome analysis of the possible regions of existence and stability of various configurations $\Delta(x)$, we summarize the results in the figure depicting the dependences $T_s(l)$, $T_a(l)$, and $T_{aa}(l)$ in the entire range of the ratios l/D. The main conclusion that can be drawn from our analysis is that the instability of the paramagnetic state relative to the formation of SDWs in the vicinity of interfaces on a scale of the order of D appears first (i.e., at a higher temperature) on the $T_s(l)$ curve corresponding to the symmetric configuration of the order parameter. This configuration exists for $T_N^0 < T < T_s(l)$ and no other configurations can appear in thin interlayers ($l < D$) in the model under investigation. In thick interlayers ($l > D$), antisymmetric configurations of type 1 (for $T_{aa}(l) < T < T_a(l)$) or type 2 (for $T_N^0 < T < T_{aa}(l)$) may also appear in addition to the symmetric configuration. The answer to the question whether the symmetric configuration be replaced by an antisymmetric one (since $T_{aa}(l) < T_a(l) < T_s(l)$ for all l) upon a decrease in temperature requires a comparison of free energies and is not obvious beforehand. Besides, it will

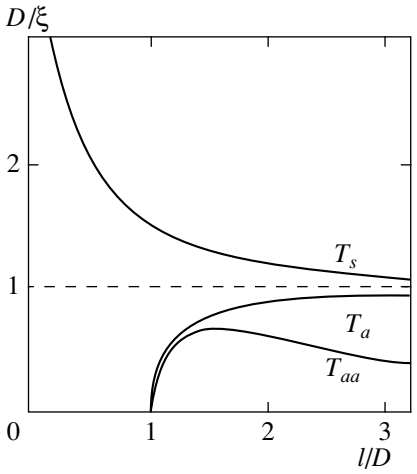

Figure.

be shown below that the free energies of the symmetric and antisymmetric (type 1) configurations become close in values in thick interlayers for $l \gg D$; in this case, in accordance with relations (19) and (20), the transition temperatures $T_s(l)$ and $T_a(l)$ asymptotically approach each other in view of the smallness of the parameter $\exp(-2l/D) \ll 1$. This means that the role of the small exchange term in the surface energy (5), which was disregarded in the above analysis, but might affect the choice of the most advantageous configuration from the energy point of view, increases considerably for $l \gg D$. Consequently, a more detailed estimation of the role of this term is required; this will be done in the next section.

4. EFFECT OF EXCHANGE INTERACTION IN THE VICINITY OF THE INTERFACE ON THE STRUCTURE OF THE SHORT-RANGE ANTIFERROMAGNETIC ORDER

The calculation of the free energy F for various SDW configurations with arbitrary relations between the quantities ξ, l, and D is a cumbersome problem. However, the situation is simplified considerably in the vicinity of the instability line $T_{s,a}(l)$, where $k^2 \ll 1$ and the expression for F can be written in a compact form suitable for obtaining qualitative estimates.

For $k'^2 \ll 1$, we obtain

$$F_\pm = -\frac{c_2 v_F^4}{4\xi^3}\sinh\left(\frac{2l}{\xi}\right) f_\pm\left(\frac{l}{\xi}\right) k_{s,a}'^4 \tag{22}$$

for the symmetric and antisymmetric structures, respectively, and the value of k' is given by expression (18). Formula (22) is valid for any ratio of lengths l and D. In

the case of a thick interlayer, when $l/D \gg 1$, we can write Eq. (22) in the form

$$F_{\pm} = -\frac{v_F}{4D}\left(\frac{\tau_0}{\tau_{s,a}(l)}\right)^2 \times\left[1 \pm 4\exp\left(-\frac{2l}{D}\right) - \left(\frac{\tau_{s,a}(l)}{\tau_0}\right)^2 \ln[\tau_{s,a}(l)]\right]^2, \quad (23)$$

where $\tau_{s,a}(l)$ are the dimensionless temperatures introduced in Section 3. It can be seen that in the limit $l/D \longrightarrow \infty$, when $\tau_a(l) \longrightarrow \tau_s(l)$, the difference between the energies of the symmetric and antisymmetric configurations is exponentially small in parameter $\exp(-2l/D) \ll 1$; i.e., these states with SDWs become almost degenerate in energy.

In a thin interlayer with $l < D$, when only the symmetric configuration exists below the transition temperature $T_s(l)$, and the antisymmetric configuration is not realized, the calculation of the free energy gives

$$F_+ = -\frac{c_2 v_F^4}{4lD^2}\left(1 - \frac{lD}{\xi^2}\right)^2. \quad (24)$$

In the vicinity of the transition temperature, dependence (24) has the form

$$F_+ \approx -\frac{v_F}{4l}\left(\frac{\tau_0}{\tau_s}\right)^2\left(1 - \frac{l}{D}\left(\frac{\tau_s}{\tau_0}\right)^2 \ln\tau_s\right)^2. \quad (25)$$

Thus, it follows from formulas (23) and (25) as well as from more general expressions for $F_{\pm}$ that the state with SDWs described by a symmetric order parameter $\Delta_+(x)$ is advantageous from the energy point of view in all regions of the (T, l) phase diagram below the transition line $T_s(l)$. However, this statement is valid only as long as we disregard the exchange term in the surface energy (5) and requires a correction when this term is taken into account. The case when the instability curves $T_s(l)$ and $T_a(l)$ converge asymptotically for $l \gg$ D so that even a slight perturbation facilitating the formation of an antisymmetric state with SDWs may change the situation radically in favor of the antisymmetric state is of special importance.

Let us consider in greater detail the explicit form of the exchange term under investigation in the case of a linearly polarized SDW. Taking into account what has been said above concerning the sign and magnitude of the exchange interaction at the Fe/Cr interface, we can write

$$F_{ex} = -\frac{A}{2}[\Delta(l) \pm (-1)^{N+1}\Delta(-l)], \quad A = J_0 S\bar{N}, \quad (26)$$

where the plus and minus signs correspond to the ferro- and antiferromagnetic relative orientation of the spins in Fe at the opposite outer layer of the three-layered structure, respectively, and N is the number of monolayers in the Cr interlayer. Let us consider the case of the ferromagnetic orientation of spins in the outer layers. If N is an odd number, the result of the inclusion of contribution (26) is almost obvious: since the symmetric state gives an additional gain in energy (26), which is equal to $F_{ex} = -A\Delta_+(l)$, it remains preferable to the antisymmetric state for any value of the ratio l/D. Besides, the exchange term in the surface energy (5), which is a source of the order parameter, leads to the emergence of a symmetric SDW for an odd N everywhere in the temperature range $T > T_s(l)$, the value of $\Delta(x)$ being small in view of the smallness of coefficient A:

$$\Delta_+^{(ind)}(x) = \frac{A\xi\cosh(x/\xi)}{2v_F^2 c_2 \sinh(l/\xi)}.$$

The order parameter induced by the magnetization of the outer layers decreases abruptly with increasing distance from the interface to the bulk of the interlayer since the length $\xi(T)$ decreases upon an increase in temperature. As $T \longrightarrow T_s(l)$, the SDW amplitude $\Delta_+(x)$ increases so that the order parameter depends on A only slightly upon a further decrease in T. Thus, the qualitative role of the exchange interaction near the interface for an odd N is mainly reduced to the formation of an induced symmetric state with a SDW in a narrow region (smaller than the interpolation length D) in the vicinity of the interface at high temperature $T > T_s(l)$ and to an additional gain in the energy of the symmetric configuration below $T_s(l)$.

The situation when N is even and the exchange term in formula (5) induces the order parameter with an antisymmetric configuration of type 1 everywhere in the temperature range $T > T_s(l)$ is much more interesting:

$$\Delta_-^{(ind)}(x) = \frac{A\xi\sinh(x/\xi)}{2v_F^2 c_2 \cosh(l/\xi)}.$$

In thin interlayers ($l < D$), when the temperature becomes slightly lower than $T_s(l)$, a symmetric state with a SDW is inevitably formed in place of the antisymmetric state $\Delta_-^{(ind)}(x)$ at a certain temperature $\tilde{T}_s(l) < T_s(l)$ for which the energy gain $|F_+|$ (24) becomes larger than $\left|F_-^{(ind)}\right| = A\Delta_-^{(ind)}(l)/2$. With such a scenario, the role of the exchange interaction at the interface at $T < T_s(l)$ is mainly reduced to a shift in the transition temperature:

$$\delta T_s = T_s - \tilde{T}_s \approx AT_s \ll T_s.$$

The possibility of the existence of a configuration with a combined symmetry below $T_s(l)$ is not considered here since the role of the component $\Delta_+(x)$ is predominant in any case.

The most interesting situation takes place in thick interlayers ($l \gg D$) for an odd number N of the monolayers, when the role of term (26) is decisive for the

choice of the symmetry of the state with a SDW below the transition point $T_s(l) \approx T_a(l)$. In the limiting case $l/D \longrightarrow \infty$, the energies $F_\pm$ (23) converge to such an extent that the exchange term (26) makes the formation of an antisymmetric state more advantageous. For $l/D \gg 1$, we can obtain the following relations:

$$F_+ - F_- = -\frac{16c_2 v_F^4}{D^3}\frac{1 - D/\xi}{1 + D/\xi} \times \left[3\left(\frac{D}{\xi}\right)^3 - 3\left(\frac{D}{\xi}\right)^2 + 1\right]\exp\left(-\frac{2l}{\xi}\right), \quad (27)$$

$$F_{\text{ex}} = -\frac{A v_F}{D}\sqrt{1 - \left(\frac{D}{\xi}\right)^2}, \quad (28)$$

which are valid in a wide temperature range $T_a(l) > T > T_{aa}(l)$. The transition line $\xi = D$ leads to the asymptotic value $T_s(l) = T_a(l) = T(\infty)$for $l/D \longrightarrow \infty$. Thus, in very thick interlayers with an even number of monolayers, the antisymmetric state with a SDW is most advantageous from the energy point of view (at least, in the vicinity of the transition temperature). As the thickness of the interlayer decreases (but still $l/D \gg 1$), the gain due to the contribution of F_{ex} becomes smaller than the gain due to the difference $F_+ - F_-$ and the conclusion that the symmetric configuration is preferred over the antisymmetric one becomes valid again for $l < \tilde{l}$. The estimate of the value of $\tilde{l}$ is given by the equality

$$\exp\left(-\frac{2\tilde{l}}{\xi}\right) = \frac{AD^2}{16c_2 v_F^3}\sqrt{\frac{(1 + D/\xi)^3}{1 - D/\xi}} \times \left[3\left(\frac{D}{\xi}\right)^3 - 3\left(\frac{D}{\xi}\right)^2 + 1\right]^{-1}, \quad (29)$$

which, however, should not be treated too rigorously since it holds only for extremely small values of $A \ll 1$ and not too close to the line $\xi = D$. The assumption that $\tilde{l}$ is larger than D only slightly (i.e., $\tilde{l}/D \geq 1$) and the difference $F_+ - F_-$ is not too small appears as more realistic. Unfortunately, the length $\tilde{l}$ can be calculated in this case only numerically, but this problem is not considered by us here.

The case of the antiferromagnetic orientation of spins of Fe in the opposite outer layers can be analyzed similarly by replacing the even number of monolayers by an odd number or vice versa. Thus, if the mutual orientation of spins in the opposite outer layers of Fe is given, either a symmetric or antisymmetric configuration with an SDW is formed depending on the number of Cr monolayers in thick interlayers ($l/D \gg 1$). The specific mechanism fixing this orientation of spins can be either an external magnetic field, or other indirect exchange channels (e.g., through electrons from the paramagnetic regions of the Fermi surface in chromium, which do not exhibit nesting and do not participate in the formation of SDWs). If, however, the external mechanisms of spin orientation in Fe are absent or insignificant for some reason, the ± sign in formula (26) is determined self-consistently depending on the number N: the plus sign should be chosen for an odd N and the minus sign for an even N. A symmetric structure of SDW is apparently most advantageous in this case in accordance with formula (26).

5. CONCLUSIONS

The main results obtained in this work can be formulated as follows.

1. Chromium-based multilayers (in particular, used in Fe/Cr type systems which have been used more comprehensively) display a considerable short- range order with a SDW initiated by the interfaces in Cr layers at temperatures much higher than the bulk transition temperature T_N. The mechanism of the formation of such a short-range order is primarily determined by the charge (and, hence, spin) density redistribution in the vicinity of the interfaces. The role of the exchange interaction itself (between the spins of band electrons of Cr and local spins of Fe in systems of the Fe/Cr type) at these interfaces is mainly reduced to the choice of the SDW structure in the situation when the symmetric and antisymmetric configurations are almost degenerate in energy. Such a situation emerges in thick chromium interlayers with $L > L^*$, where L^* is the characteristic thickness consisting of 20–30 monolayers.

2. The characteristic scale of the short-range order emerging in the chromium interlayer below a certain temperature $T_1(L) > T_N(L)$ is the so-called interpolation length which is mainly determined by the parameters of the interlayer material and the effective surface potential at the interfaces as well as (to a considerably smaller extent) by the temperature. The temperature $T_1(L)$ increases considerably in thin interlayers ($L < 2D$) and attains a constant value independent of L in thick interlayers ($L > 2D$). The length D is larger than the separation between chromium monolayers; in the present work, we can assume the estimate $2D \approx L^*$ which makes it possible to carry out a qualitative comparison with some experimental results.

Let us discuss two groups of reliable experimental results which provide information on the magnetic structure in the chromium interlayer. We consider first some results on the exchange interaction between the ferromagnetic outer layers in the triple layer Fe/Cr/Fe in which the chromium interlayer has the shape of a wedge. Important results were obtained in a series of publications (all references can be found in review [1]) devoted to an analysis of the so-called optimized triple Fe/Cr/Fe layers formed by the layer-by-layer deposition of chromium on a thick and almost perfectly planar iron (001) substrate having the temperature $T_{\text{opt}} \approx 570 \pm 20$ K. The spatial configuration of the chromium layer

in the growth front resembles wide terraces between monatomic steps with a mean size of several tens of nanometers. The method of scanning tunnel microscopy with polarization analysis makes it possible to detect, apart from the expected alternation of ferro- and antiferromagnetic links in the Fe layers upon the change in the number N of chromium monolayers by unity, an indirect exchange phase slip at room temperature for $N \approx 24$, 44, and 64 [19]. It is generally accepted that this fact indirectly indicates the presence of a transverse SDW in the Cr interlayer with an amplitude modulation along the direction of the heterostructure growth, the modulation preserving its sign for thicknesses with $N < N_1 \approx 24$. It was found that the temperature dependence of N_1 is preserved up to temperatures $T \approx 550$ K, at which the value of N_1 becomes as high as $N_1(550\ \text{K}) \approx 38$; i.e., the antiferromagnetic order exists in the Cr interlayer at temperatures much higher than the bulk transition temperature $T_N = 311$ K. In the model proposed by us, the SDW contribution to the exchange interaction leads to a phase slip for $L = \tilde{L}$, where $\tilde{L}$ is the thickness corresponding to the change in the SDW configuration from the symmetric to the antisymmetric one. The estimate $\tilde{L} \approx (N_1 + 1)a/2$, where $a/2$ is the separation between monolayers, gives $\tilde{L} \approx 55$ Å for the high-temperature phase. Traditionally, the phase slip in the indirect exchange is attributed to the periodic SDW modulation in chromium, but this assumption is far from obvious in the temperature range above T_N we are dealing with. We proved that even in the absence of a periodic modulation (incommensurability), the existence of at least the first critical number N_1 of monolayers below which the phase slip is absent can be explained at a high temperature $T > T_N$, when the long-range order and incommensurability are almost ruled out. Unfortunately, the temperature dependences of the second and higher critical number of monolayers (N_2, N_3, …) are unknown to us, although the emergence of these numbers at $T > T_N$ would indicate the existence of more complex structures with SDWs than those determined above; this does not contradict, in principle, the general concept used by us here.

In contrast to triple Fe/Cr/Fe layers, the study of superlattices with tens and hundreds of periods makes it possible to analyze the state with an SDW directly using the sensitive neutron-diffraction methods. The magnetic phase diagram of structures of the Fe/Cr type was constructed thanks mainly to the efforts of two groups of experimenters [8–10, 20]. Among other things, the critical thickness $L = L^*$ of the chromium interlayer for which the traditional antiferromagnetic ordering with an incommensurate SDW observed in bulk chromium ($L^* \approx 45$ Å) disappears was established [8]. For $L < L^*$, the Néel temperature $T_N(L)$ is formally equal to zero, while for $L > L^*$ it increases sharply and rapidly tends to the "bulk" value $T_N(\infty) = 311$ K. The pattern of magnetic reflections in neutron scattering in the form of a central peak with two satellites, which is observed everywhere below $T_N(L)$, is successfully interpreted in terms of an incommensurate SDW with a large period which virtually coincides with the corresponding period in bulk chromium [9]. However, magnetic reflections in the form of a central peak and corresponding (at first glance) to the existence of a commensurate SDW are also observed above $T_N(L)$ for $L > L^*$ as well as in multilayers with $L < L^*$. These reflections are preserved up to temperatures $T_1 \approx 550$–600 K [10, 20]. In our model, such a "high-temperature" SDW phase is naturally explained as a state with a short-range order emerging below $T_1(L)$ and having the characteristic scale $D \approx L^*/2$. Thus, only a part of the interlayer of thickness approximately equal to $2D$ displays a clearly manifested antiferromagnetic order in the temperature range $T_N(L) < T < T_1(L)$, while the middle of the interlayer of thickness $l_{\text{eff}} = L - 2D$ essentially remains almost paramagnetic. We can assume that it is precisely the thickness L_{eff} that appears in the scaling dependence $T_N(l_{\text{eff}})$ [8] discussed in the Introduction, although this dependence has not been calculated directly as yet.

In this paper, many important problems listed below are not considered.

(i) The possibility of formation of noncollinear SDW structures below $T_1(L)$ in the range of applicability of the Ginzburg–Landau expansion (3)–(6) for the free energy $F(\mathbf{\Delta})$.

(ii) The determination of the SDW structure in the range of temperatures close to $T_N(L)$. It should be recalled that a decrease in temperature leads to the sign reversal of coefficient c_2 in expansion (5) and necessitates the inclusion of higher-order gradient terms in this expansion. The solutions obtained by us for $\mathbf{\Delta}(x)$ are inapplicable even at $T \geq T_N(L)$, i.e., in the short-range order region, if $c_2(T) < 0$.

(iii) The determination of the structure of the antiferromagnetic order with SDW in the low-temperature range $T \ll T_N(L)$. Unfortunately, we cannot offer any approach other than the variational approach for analyzing this problem in view of the considerable difficulties involved in the solution of the self-consistent equation for the order parameter $\mathbf{\Delta}(x)$.

The last thing which we will touch upon in this paper is the need to expand the class of Cr-based multistructures under investigation. It would be very interesting to obtain neutron-diffraction data for a structure of the Co/Cr type in which the exchange coupling at the interface is much stronger than in Fe/Cr structures and is apparently ferromagnetic. It would also be interesting to study the $Fe/Cr_{1-x}Mn_x$ type structures for low (up to 1%) concentrations of manganese since such structures must display a sharp change in the SDW type and an increase in the transition temperature T_N. Systems of the $Fe/Cr_{1-x}Fe_x$ type [21] with a low (up to 3–

5%) concentration of Fe, in which a change in the SDW structure is accompanied by a small decrease in T_N, may turn out to be promising for an analysis of fine features of the short-order mechanism with SDWs. This list can be extended, but even the above examples indicate, in our opinion, the possible direction of future investigations.

ACKNOWLEDGMENTS

One of the authors (V. V. T) is grateful to B. A. Volkov, Yu. A. Uspenskiĭ, and other participants of the seminar on solid-state physics at the Theoretical department of Lebedev Physical Institute, Russian Academy of Sciences for fruitful discussions of the results of this work.

This research was partially supported 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. S. Parkin, N. More, and K. Roche, Phys. Rev. Lett. **64**, 2304 (1990).
4. E. Fawcett, H. L. Alberts, V. Yu. Galkin, *et al.*, Rev. Mod. Phys. **66**, 25 (1994).
5. T. M. Rice, Phys. Rev. B **2**, 3619 (1970).
6. K. Machida and M. Fujita, Phys. Rev. B **30**, 5284 (1984).
7. N. I. Kulikov and V. V. Tugushev, Usp. Fiz. Nauk **144**, 643 (1984) [Sov. Phys. Usp. **27**, 954 (1984)].
8. E. E. Fullerton, K. T. Riggs, C. H. Sowers, and S. D. Bader, Phys. Rev. Lett. **75**, 330 (1995).
9. E. E. Fullerton, S. D. Bader, and J. L. Robertson, Phys. Rev. Lett. **77**, 1382 (1996).
10. A. Schreyer, C. F. Majkrzak, T. Zeidler, *et al.*, Phys. Rev. Lett. **79**, 4914 (1997).
11. K. Binder, in *Phase Transitions and Critical Phenomena*, Ed. by C. Domb and J. L. Lebowitz (Academic, London, 1983), p. 1.
12. M. Avignon, V. Men'shov, and V. Tugushev, in *Proceedings of JEMS01, Grenoble, France, 2001*.
13. R. S. Fishman and Z. P. Shi, Phys. Rev. B **59**, 13849 (1999); R. S. Fishman, Phys. Rev. Lett. **78**, 1351 (1997).
14. A. M. Niklasson, B. Johansson, and L. Nordstrom, Phys. Rev. Lett. **82**, 4544 (1999).
15. V. V. Tugushev, in *Modern Problems in Condensed Matter Sciences*, Vol. 32: *Electronic Phase Transitions*, Ed. by W. Hanke and Yu. V. Kopaev (North-Holland, Amsterdam, 1992), p. 239.
16. A. I. Buzdin, V. N. Men'shov, and V. V. Tugushev, Zh. Éksp. Teor. Fiz. **91**, 2204 (1986).
17. E. I. Kondorskiĭ, T. I. Kostina, and V. P. Medvedchikov, Vestn. Mosk. Univ., Ser. 3: Fiz., Astron. **36**, 22 (1981).
18. *Handbook of Mathematical Functions*, Ed. by M. Abramowitz and I. A. Stegun (Dover, New York, 1971; Nauka, Moscow, 1979).
19. J. Unguris, R. J. Celotta, and D. T. Pierce, Phys. Rev. Lett. **69**, 1125 (1992).
20. A. Schreyer, J. F. Ankner, T. Zeidler, *et al.*, Phys. Rev. B **52**, 16066 (1995).
21. E. E. Fullerton, C. H. Sowers, and S. D. Bader, Phys. Rev. B **56**, 5468 (1997).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 93, No. 4, 2001, pp. 796–804.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 910–920.

SOLIDS
Structure

Magnetic Phase Transitions and Phase Diagrams of $DyMn_2Ge_2$

Guo Guanghua[a], M. V. Eremin[b], A. Kirste[c], N. P. Kolmakova[b, **], A. S. Lagutin[d], R. Z. Levitin[a, *], M. von Ortenberg [c], and A. A. Sidorenko[b]

[a] *Moscow State University, Vorob'evy gory, Moscow, 119899 Russia*
**e-mail: levitin@plms.phys.msu.su*
[b] *Bryansk State Technical University, Bryansk, 241035 Russia*
***e-mail: npk@bitmcnit.bryansk.su*
[c] *von Humboldt Universität, D-10115, Berlin, Germany*
[d] *Institute of Molecular Physics, Russian Research Centre, Kurchatov Institute, pl. Kurchatova 1, Moscow, 123182 Russia*

Received April 24, 2001

Abstract—The experimental studies of magnetic phase transitions in the layered tetragonal intermetallic compound $DyMn_2Ge_2$ are continued. The existence of spontaneous phase transitions is confirmed by the results of measurements of the temperature dependences of lattice parameters and the initial magnetic susceptibility. The measurements in strong (up to 50 T) and ultrastrong (up to 150 T) fields revealed two new field-induced magnetic transitions. The inclusion of the exchange interaction between next-to-nearest layers of manganese and the crystal field effects for the rare-earth subsystem along with the antiferromagnetic exchange interaction of nearest Mn layers has made it possible to describe the magnetic properties of $DyMn_2Ge_2$ in a wide range of magnetic fields. The parameters of these interactions are determined from a comparison of the experimental and theoretical magnetization curves and *H–T* phase diagrams. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The magnetic and other physical properties of ternary intermetallic compounds RMn_2Ge_2 (R stands for rare-earth elements) are of considerable interest due to the effects associated with the co-existence of two magnetic subsystems: the band 3*d* (manganese) subsystem and the localized 4*f* (rare-earth) subsystem. Besides, these compounds are perfect natural superlattices which may display the effect of giant magnetoresistance, e.g., during the magnetic field induced metamagnetic transitions. The layered structure, the high sensitivity of exchange parameters to atomic spacing, the antiferromagnetic exchange interaction in the Mn subsystem in intermetallic compounds with heavy rare-earths, and pronounced crystal field effects in the rare-earth subsystem lead to complex and very interesting magnetic phase diagrams of these compounds.

The RMn_2Ge_2 compounds crystallize in the a tetragonal structure of the $ThCr_2Si_2$ type (space group *I*4/*mmm*), which has the form of a set of R–Ge–Mn–Ge–R perpendicular to the *c* axis. The magnetic properties of these compounds were intensely studied during the last two decades [1–4]; it was found that numerous magnetic phase transitions take place in them. In these intermetallic compounds, both rare-earth and manganese possess magnetic moments. The Mn–Mn exchange interaction in a layer is the strongest of all interactions; it determines the ferromagnetic ordering of the magnetic moments of Mn in the layer. In the case of heavy rare-earths, the exchange interactions between Mn–Mn and R–Mn layers are antiferromagnetic. At room temperature, the Mn subsystem is antiferromagnetically ordered, while the rare-earth subsystem is disordered since the effective fields exerted on this subsystem by two adjacent Mn layers compensate each other, and the intrinsic exchange in the rare-earth subsystem is small. The magnetic moments of the rare earth become ordered only at low temperatures. For example, $GdMn_2Ge_2$ and $TbMn_2Ge_2$ compounds experience a first-order transition to a state with an ordered rare earth at T = 96 and 95 K, respectively [2]. Below this temperature, the R–Mn exchange interaction induces the formation of a collinear ferrimagnetic structure: the magnetic moments of the rare earth are oriented along the tetragonal axis, while the magnetic moments of all Mn atoms are antiparallel to them.

The magnetic behavior of $DyMn_2Ge_2$ is more intricate. The magnetic and neutron-diffraction studies on single crystals [5] proved that the magnetic moments of Mn in $DyMn_2Ge_2$ become antiferromagnetically ordered at T_N = 431 K. According to the results obtained in [5] and in [6, 7], where neutron-diffraction experiments were made on powders, $DyMn_2Ge_2$ experiences two first-order magnetic phase transitions in the low-temperature region: at temperatures T_1, which lie in the interval from 33 to 35 K, and at T_2 varying between

37.5 and 40 K. At $T < T_1$, this compound is characterized by a collinear ferrimagnetic (Fi) structure similar to the structure of $GdMn_2Ge_2$ and $TbMn_2Ge_2$ described above. At $T > T_2$, the compound has an antiferromagnetic (AF) structure with disordered moments of dysprosium. For the temperature range between T_1 and T_2, the results presented in [5–7] are contradictory. According to [6], three phases coexist in this temperature range: Fi, AF, and an intermediate phase in which all magnetic moments are also directed along the *c* axis, the directions of the magnetic moments of Mn in the layers alternate in the sequence ↓↑↓↓↑↓..., and the magnetic unit cell is characterized by tripling along the tetragonal axis, i.e., $a' = a$ and $c' = 3c$. According to [7], the AF phase does not exist below T_2, which follows from the temperature dependence of the integrated intensity of the corresponding line. The authors of [7] assume that between T_1 and T_2, there exists a single incommensurate phase whose wave vector was determined preliminarily as $\mathbf{k} \approx (0, 0, 0.65)$. The measurements of the magnetization curves on $DyMn_2Ge_2$ single crystals in fields up to 15 T [6] and 5 T [8] revealed the existence of first-order phase transitions in all temperature intervals under investigation up to 70 K. No attempts were made to describe theoretically the magnetic phase diagrams of $DyMn_2Ge_2$ except an attempt in [6] to determine the exchange parameters from the condition for the existence of spontaneous first-order transition from the ferrimagnetic to the antiferromagnetic phase in the vicinity of 40 K, which was made not quite correctly.

This work is devoted to further experimental investigations of magnetic properties of $DyMn_2Ge_2$ and the theoretical description of magnetic phase transitions at different temperatures and for different magnetic field directions, the construction of the *H–T* phase diagrams, and the determination of the parameters of interactions in this compound on the basis of the model proposed by us here.

2. EXPERIMENTAL RESULTS

Polycrystalline samples of the intermetallic compound $DyMn_2Ge_2$ were melted in an induction oven in argon under the quasilevitation conditions from the initial elements of purity 99.9. To achieve a high degree of homogeneity, the samples were melted thrice and annealed for 170 h in a dynamic vacuum at 750°C. The presence of only one phase in the sample was checked with the help of X-ray diffraction.

The initial magnetic susceptibility was measured both in a varying field (in the temperature range from 4.2 to 270 K) and in a constant field (at temperatures from 300 to 500 K). Such measurements make it possible to reliably determine the singularities on the temperature dependence of magnetic susceptibility emerging during magnetic phase transitions. The temperature dependence of susceptibility, $\chi(T)$, clearly exhibits

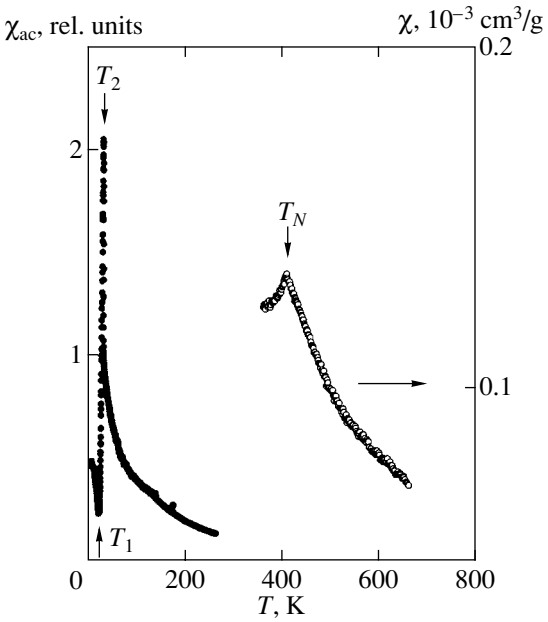

Fig. 1. Temperature dependence of the initial magnetic susceptibility of $DyMn_2Ge_2$, measured in a varying field (dark circles) and static field (light circles): T_1 and T_2 are the temperatures of spontaneous magnetic phase transitions (see text), T_N is the magnetic ordering temperature.

anomalies (Fig. 1) associated with low-temperature phase transitions as well as an anomaly in the transition from the paramagnetic to the antiferromagnetic state in the Mn subsystem at $T_N \approx 440$ K.

The temperature dependence of the lattice parameters was measured by the X-ray diffraction on a "Geigerflex" diffractometer (Japan) in the temperature range 10–800 K. The temperature dependences of the lattice parameters of $DyMn_2Ge_2$ presented in Fig. 2 clearly show that the transition from the paramagnetic to the antiferromagnetic state in the Mn subsystem ($T_N \approx 440$ K) is accompanied by a noticeable anomaly on the $a(T)$ dependence; i.e., the temperature dependence of the Mn–Mn atomic spacing in the layer changes during this transition. Low-temperature transitions are manifested on the temperature dependences less strongly for reasons which are not completely clear. This is probably due to the fact that, according to [6], different magnetic structure coexist in $DyMn_2Ge_2$ at low temperature and, hence, the X-ray reflections corresponding to different magnetic phases with different lattice parameters cannot be separated. This hypothesis is supported by the fact that, according to the results of our measurements, the X-ray reflection (220) is broadened significantly below T_2.

The magnetization and the field dependence of dM/dt were measured in strong pulsed magnetic fields up to 50 T with a pulse duration of 26 ms with the help of a pulsed coil magnetometer [9] in the temperature range from 7 to 70 K on free powders whose particles can rotate in a field. Figure 3 shows the field depen-

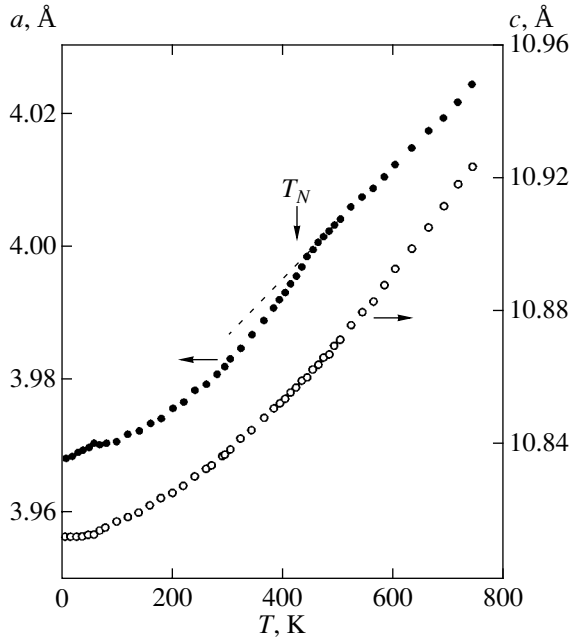

Fig. 2. Temperature dependences of the crystal lattice parameters for $DyMn_2Ge_2$.

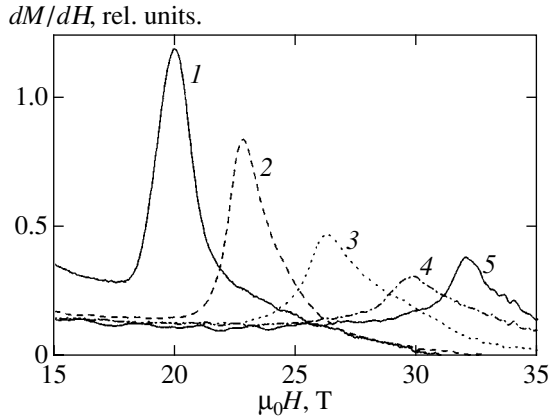

Fig. 3. Differential magnetic susceptibility of $DyMn_2Ge_2$, measured on free powders in fields up to 50 T at the following temperatures, K: 58 (*1*), 48 (*2*), 32 (*3*), 18 (*4*), and 7 (*5*). The curves are recorded in increasing fields. The curves recorded in decreasing fields are displaced towards lower fields approximately by 1 T.

dences of the differential magnetic susceptibility dM/dH at various temperatures, which were measured in an increasing field. A first-order phase transition corresponding to the peak on the curve can be seen clearly. The field dependences of susceptibility recorded in a decreasing field are similar to those presented in Fig. 3 for an increasing field, but the peaks are displaced towards lower fields approximately by 1 T, indicating the existence of a hysteresis during the phase transition.

In pulsed magnetic fields up to 150 T, which are generated during the discharge of a capacitor bank through a one-turn solenoid (the duration of the first half-period of the pulse is 6 μs), the measurements were made by the induction technique on fixed powders with a particle size of about 10 μm at temperatures from 5 to 7 K. Using this method, it is impossible to compensate the measuring coils completely and the signal is described by the formula

$$V = (a + b\chi)\frac{dH}{dt}.$$

The first term in the parentheses is the signal associated with the decompensation of the coils, the second is the signal from the sample, χ is the differential magnetic susceptibility of the sample, and a and b are the constants associated with the parameters of the coils and the sample, respectively (see [10] for details). In the critical field of the first-order phase transition, susceptibility χ passes through a peak, which must be reflected in the emergence of a peak in the voltage V induced in the measuring coils. It should be noted that the fields H_{max} at which the peaks on the $V(H)$ and $\chi(H)$ dependences are observed coincide exactly only when dH/dt = const, which was indeed the case during measurements in the fields up to 50 T. If the transition takes place for $H > 0.5H_{max}$, the value of the critical field must be determined from the field dependence of the quantity $V/(dH/dt)$, which is just what we did in the present case.

The obtained dependence of the differential magnetic susceptibility of $DyMn_2Ge_2$ is shown in Fig. 4. Distinct peaks of dM/dH correspond to a phase transition. The values of critical fields upon an increase (112 T) and a decrease (108 T) of the field are different, which is typical of first-order phase transitions. Besides, the relaxation effects associated with the pulsed nature of magnetization in strong fields apparently affect the width of the hysteresis loop also.

3. THEORETICAL APPROACH

In our previous publications, the experimental data for intermetallic compounds $Gd_{1-x}Y_xMn_2Ge_2$ and $Gd_{1-x}La_xMn_2Ge_2$ [11–13] were described by using a model in which the interactions between the corresponding nearest layers of magnetic atoms were taken into consideration along with the exchange interactions in rare-earth and Mn layers in the molecular field approximation. This model made it possible to correctly describe the phase transitions observed in these intermetallic compounds, but in some cases the experimental and theoretical data could not be matched quantitatively. An analysis of experimental results for $DyMn_2Ge_2$ shows that this simple model does not allow to describe the magnetic properties of this intermetallic compound even qualitatively. In particular, the possibility of interpreting the existence of a magnetic structure observed in the temperature range between T_1 and T_2, the magnetization jump during the first-order phase

transition at low temperatures in a field oriented along the tetragonal axis of the crystal (see [5]), etc. cannot be explained in this theory. For this reason, we complicated the model by taking into account the exchange interactions between next-to-nearest magnetic layers. This appears quite natural since, apart from indirect exchange through germanium, the exchange interaction in the intermetallic compounds under investigation occurs through the conduction electrons and is a long-range interaction. Besides, in contrast to the Gd^{3+} ion, the Dy^{3+} ion is not an *S* ion and, hence the effects associated with the crystal field are significant in this case. When these factors are taken into account, nonequivalent positions of magnetic moments may appear both in the dysprosium and in the manganese magnetic subsystems of the crystal. The existence of such nonequivalent positions and the exchange interactions between different layers of magnetic atoms were precisely taken into consideration while writing the expressions for the effective Hamiltonians.

The effective Hamiltonian for a Dy^{3+} ion in the *i*th position in the molecular field approximation can be written in the form

$$\mathcal{H}_{\mathrm{Dy}}^{(i)} = B_2^0 O_2^0 + B_4^0 O_4^0 + B_4^4 O_4^4 + B_6^0 O_6^0 + B_6^4 O_6^4 - g_J \mu_{\mathrm{B}} \mathbf{J}^{(i)} \cdot (H + H_m^{(i)}), \quad (1)$$

where B_n^m are the crystal field parameters of the tetragonal symmetry, O_n^m are the equivalent operators, g_J is the Lande factor, and $\mathbf{J}^{(i)}$ is the angular momentum operator of the Dy^{3+} ion. The molecular field strength $H_m^{(i)}$ is defined as

$$H_{mj}^{(i)} = \sum_l \lambda_{11}^{(l)} M_j^{(l)} + \sum_k \lambda_{12}^{(k)} m_j^{(k)}, \quad j = x, y, z. \quad (2)$$

The components of the *i*th dysprosium ($M^{(i)}$) and of the *k*th manganese ($m^{(k)}$) magnetic moments are given by

$$M_j^{(i)} = \mu_{\mathrm{B}} g_J \langle J_j^{(i)} \rangle, \quad m_j^{(k)} = \mu_{\mathrm{B}} g \langle S_j^{(k)} \rangle,$$

where g is the g-factor of Mn, $\mathbf{S}^{(k)}$ is the spin moment operator for Mn, and $\lambda_{11}^{(l)}$ and $\lambda_{12}^{(k)}$ are the parameters of Dy–Dy and Dy–Mn exchange interactions, respectively. It should be noted at the very outset that, according to our calculations, the parameters of the exchange interaction between Dy–Dy and Dy–Mn atoms corresponding to different layers appear additively in all expressions; for this reason, we will subsequently use the total parameters λ_{11} and λ_{12}.

For itinerant electrons of the manganese *d* subsystem, we will take into account only exchange interactions in the effective Hamiltonian $\mathcal{H}_{\mathrm{Mn}}^{(k)}$ for the *k*th Mn atom:

$$\mathcal{H}_{\mathrm{Mn}}^{(k)} = -g\mu_{\mathrm{B}} \mathbf{S}^{(k)} \cdot \mathbf{H}_{\mathrm{Mn}}^k. \quad (3)$$

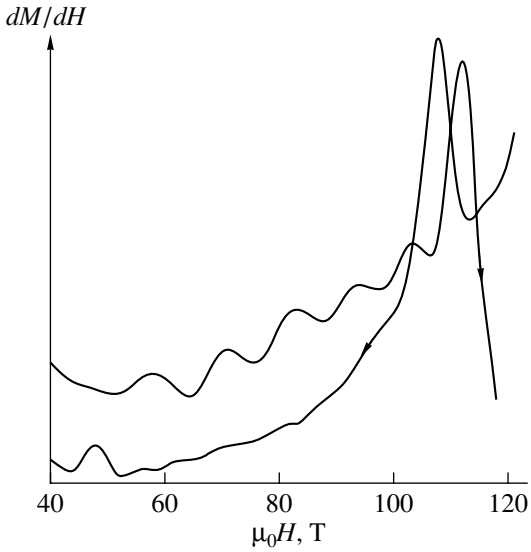

Fig. 4. Differential magnetic susceptibility at $T = 5$ K, measured in pulsed fields up to 150 T. Arrows distinguish the curves obtained upon the application and removal of the field.

The effective field acting on the *k*th Mn atom in the external magnetic field **H** oriented at an angle φ to the *c* axis of the crystal is given by

$$H_{\mathrm{Mn}}^{(k)} = H\cos(\varphi - \eta_k) + H_m^{(k)},$$
$$H_m^{(k)} = \sum_{n = k,\, k \pm 1,\, \ldots} \lambda_{22}^{(n)} m^{(n)} \cos(\eta_n - \eta_k) + \lambda_{12} \sum_i (M_z^{(i)} \cos\eta_k + M_x^{(i)} \sin\eta_k), \quad (4)$$

where η_k is the polar angle of the *k*th magnetic moment of Mn and $\lambda_{22}^{(n)}$ are the parameters of exchange interaction between Mn atoms belonging to the *k*th and *n*th layers. The anisotropy of the manganese subsystem will be included in the thermodynamic potential as an additive term in view of its smallness in comparison with the exchange. It should be noted that, for the convenience of notation, we have included, in addition to the exchange constant *I*, the factors required for writing the effective Hamiltonians and molecular fields in the form (1)–(4) in the exchange parameters λ introduced above; namely,

$$\lambda_{11} = \frac{(g_J - 1)^2}{g_J \mu_{\mathrm{B}}} I_{11}, \quad \lambda_{12} = \frac{g_J - 1}{g_J \mu_{\mathrm{B}}} I_{12},$$
$$\lambda_{22}^{(n)} = \frac{1}{g\mu_{\mathrm{B}}} I_{22}^{(n)}.$$

The thermodynamic potential per formula unit in the molecular field approximation is defined as

$$\Phi = \frac{1}{N}\Bigg\{-k_BT\sum_{i=1}^{N}\ln Z_i + \frac{1}{2}\sum_{i=1}^{N}\mathbf{M}^{(i)}\cdot H_m^{(i)} - k_BT\sum_{i=1}^{2N}\ln\frac{\sinh[(2S+1)x_k/2]}{\sinh(x_k/2)} + \frac{1}{2}\sum_{k=1}^{2N}\mathbf{m}^{(k)}\cdot H_m^{(k)} + \sum_{k=1}^{2N}K\sin^2\theta_k\Bigg\} \quad (5)$$

(N *i*s the number of nonequivalent formula units). The partition function Z_i for the ith dysprosium moment was evaluated using numerical diagonalization of the Hamiltonian $\mathcal{H}_{\mathrm{Dy}}^{(i)}$ (1), involving the solution of the corresponding self-consistent problems, $x_k = \mu_B g H_{\mathrm{Mn}}^{(k)}/k_BT$, and K is the anisotropy constant for the Mn subsystem. The second and fourth terms in formula (5) are conventional correcting terms in the molecular field theory.

For known values of the parameters of the system appearing in the thermodynamic potential, we can find the range of existence of various magnetic structures from the conditions of the minimum of their thermodynamic potential, and calculate the phase-transition fields (from the equality of the thermodynamic potentials of different phases in the case of first-order phase transitions). In the present work, we will solve the inverse problem of determining the complete set of parameters of $DyMn_2Ge_2$ for interpreting the experimental data concerning the magnetic properties of this compound on the basis of thermodynamic potential (5).

4. PHASE TRANSITIONS IN A FIELD PARALLEL TO THE TETRAGONAL AXIS

Low-temperature transitions. We begin our analysis with the phase transitions induced by a magnetic field parallel to the tetragonal axis at helium temperatures. It is known [5] that at this temperature, the initial phase is the ferrimagnetic phase (Fi) with the magnetic moments of Mn in all the layers oriented antiparallel to moments $\mathbf{M}_{\mathrm{Dy}}$ and that a first-order phase transition occurs in the vicinity of 7 T. According to our experimental data, two more first-order phase transitions take place in stronger fields: near 30 T (see Fig. 3) and near 110 T (see Fig. 4). The first-order phase transition near 110 T was detected during measurements on powder samples. We attribute it to crystallites for which the field is oriented along the tetragonal axis since out calculations revealed that no other first-order phase transitions are possible in this field regions for other mutual orientations of the field and the tetragonal axis. From the value of magnetization after the jump at $H \approx 7$ T [5], it can be concluded that the transition occurs to a phase (which will be referred as the intermediate phase I) in which the magnetic moment of every third layer of the Mn subsystem is reoriented along the direction of the field.[1] The spontaneous formation of precisely this phase was presumed in neutron-diffraction and Mössbauer studies in the temperature range between T_1 and T_2 [6]. This structure contains antiferromagnetic and ferromagnetic Mn blocks (see Fig. 5 in [6]) so that the entire phase I is in fact an ordered superposition of the Fi and the antiferromagnetic (AF) phases, in which a block of the Fi phase follows two blocks of the AF phase. Obviously, there are two types of Dy atoms in this case. In zero field, dysprosium in the Fi blocks is magnetized due to the Dy–Mn exchange, while in the AF blocks, Dy is disordered. It was proposed in [5] that the transition near 7 T occurs into the triangular phase in which the magnetic moment of Dy is oriented, as before, along the tetragonal axis (and along the field), while the moments of manganese form obtuse angles with $\mathbf{M}_{\mathrm{Dy}}$. This assumption contradicts the experimental data obtained by the same authors [5], according to which the susceptibility of the system is virtually equal to zero in fields above the transition field, while its magnitude in the triangular phase must be significant.

If the Fi $\longrightarrow$ I transition takes place in the vicinity of 7 T, we can assume that the antiferromagnetic (AF) phase with the antiferromagnetic ordering in the Mn subsystem and the ferromagnetic (F) phase in which the magnetic moments of both subsystems are oriented along the field play the role of higher-field phases. The values of the three critical fields of first-order phase transitions at helium temperatures,

$$H_{\mathrm{Fi}\to\mathbf{I}} = 7\ \mathrm{T},\quad H_{\mathbf{I}\to\mathrm{AF}} = 32\ \mathrm{T},\quad H_{\mathrm{AF}\to\mathrm{F}} = 110\ \mathrm{T},$$

make it possible to determine three exchange parameters of the compound: λ_{12}(Dy–Mn), $\lambda_{22}^{(2)}$ (Mn–Mn interaction in adjacent layers), and $\lambda_{22}^{(3)}$ (Mn–Mn interaction in alternate layers). Since we are dealing with the situation when temperatures are low, the external field is parallel to the tetragonal axis and is obviously too high for all the magnetic moments to be saturated ($M_{\mathrm{Dy}} = M$, $m_1 = m_2 = \ldots = m$), we can use the expres-

[1] It should be noted that the magnitude of magnetization after the phase transitions suggests the reorientation of the magnetic moment of every fourth Mn layer. However, our calculations proved that the energy of such a state is higher, while the phase-transition fields coincide in the approximation taking into account the exchange interactions between Mn layers separated by not more than one layer. The inclusion of the antiferromagnetic Mn–Mn exchange interaction between the layers separated by two layers removes the degeneracy of the critical fields, the critical field $H_{\mathrm{Fi}\to\mathrm{I}}$ being characterized by a smaller value.

sions for the critical fields obtained from the low-temperature expansion of the thermodynamic potential (5):

$$\begin{aligned} H_{\mathrm{Fi}\to\mathrm{I}} &= -\lambda_{12}M + (\lambda_{22}^{(2)} + \lambda_{22}^{(3)})m, \\ H_{\mathrm{I}\to\mathrm{AF}} &= -\lambda_{12}M + (\lambda_{22}^{(2)} - 2\lambda_{22}^{(3)})m, \\ H_{\mathrm{AF}\to\mathrm{F}} &= -\lambda_{12}M - (\lambda_{22}^{(2)} + \lambda_{22}^{(3)})m. \end{aligned} \tag{6}$$

At helium temperatures, the magnetic moment of Dy is formed, to a high degree of accuracy, by the lower level of the ground Kramers doublet split by the exchange and external magnetic fields. According to the results obtained in [6], this level is a nearly pure $|15/2\rangle$ state; hence we put $M = 10\mu_B$. Assuming that $m = 2.2\mu_B$ in accordance with [5], we find that

$$\begin{aligned} \lambda_{12} &= -5.85\frac{\mathrm{T}}{\mu_B}, \quad \lambda_{22}^{(2)} = -20\frac{\mathrm{T}}{\mu_B}, \\ \lambda_{22}^{(3)} &= -3.5\frac{\mathrm{T}}{\mu_B}. \end{aligned} \tag{7}$$

The total low-temperature magnetization curve of the compound $DyMn_2Ge_2$ calculated with these parameters for a field directed along the tetragonal axis is shown in Fig. 5. The same figure shows the experimental curve measured in a field up to 15 T in [5] as well as high-field peaks of differential magnetic susceptibility obtained by us here. It should be noted that for the values (7) of exchange parameters, the triangular phase cannot emerge in a field along the tetragonal axis.

The value of the Mn–Mn exchange interaction parameter in a layer can be determined from the value of $T_N = 440$ K with known $\lambda_{22}^{(2)}$ and $\lambda_{22}^{(3)}$:

$$\lambda_{22}^{(1)} = 2\times 10^3\frac{\mathrm{T}}{\mu_B}.$$

It is well known that the Mn–Mn exchange interaction between the layers in layered compounds RMn_2Ge_2 is very sensitive to the atomic spacing in a layer, i.e., to the lattice parameter a. This leads to the temperature dependence of $\lambda_{22}^{(2)}$ associated with thermal expansion in the form [14]

$$\lambda_{22}^{(2)}(T) = \rho[a(T) - a_c], \tag{8}$$

where $a_c = 4.045$ Å for intermetallic compounds RMn_2Ge_2. However, our calculations proved that in the temperature interval below 70 K investigated by us here, the inclusion of the temperature dependence of $\lambda_{22}^{(2)}$ in $DyMn_2Ge_2$ is not necessary.

***H–T* diagram.** The temperature dependences of critical fields must be calculated on the basis of the general expression (5) for thermodynamic potential. The crystal field parameters for the Dy^{3+} ion in $DyMn_2Ge_2$ were determined in [6] from the temperature dependences of the hyperfine field and quadrupole interaction, which were obtained from an analysis of the Mössbauer spectra.The sixth-order parameters B_6^0 and B_6^4 were put equal to zero, while the remaining parameters were determined to within 30%. It will be shown below that in order to describe the processes of magnetization in a field perpendicular to the tetragonal axis, the values of some parameters from those given in [6] have to be changed within the above-indicated error. For **H** || **c** the crystal field effects and, hence, the variations of parameters are less significant. It is impossible to determine unambiguously the parameter λ_{11} of the Dy–Dy exchange from the available experimental data, but the absence of the Fi phase above T_2 sets the upper limit on its value. We take it equal to $0.49\mathrm{T}/\mu_B$; the Curie temperature of the Dy subsystem corresponding to this value of λ_{11} amounts to 25 K.

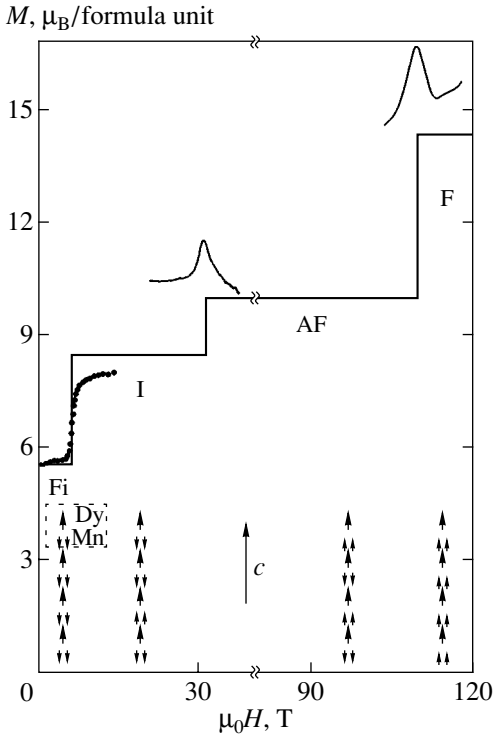

Fig. 5. Low-temperature magnetization of $DyMn_2Ge_2$ calculated per formula unit as a function of the field applied along the tetragonal axis. The curves correspond to calculations and dark circles correspond to the experimental data obtained for a single crystal at 4.2 K [5]. The *dM*/*dH* peaks shown above the magnetization curve for the I ⟶ AF and AF ⟶ F phase transitions were obtained by us at $T = 7$ and 5 K, respectively. The arrows indicate the magnetic moments of dysprosium and manganese in the layers. The formula unit is shown in the dashed frame.

The theoretical *H–T* phase diagram is shown in Fig. 6 together with the experimental data obtained in [5, 8] and in the present work. It should be emphasized that all phase transitions are of the first order and the corresponding magnetization curves obtained as a result of

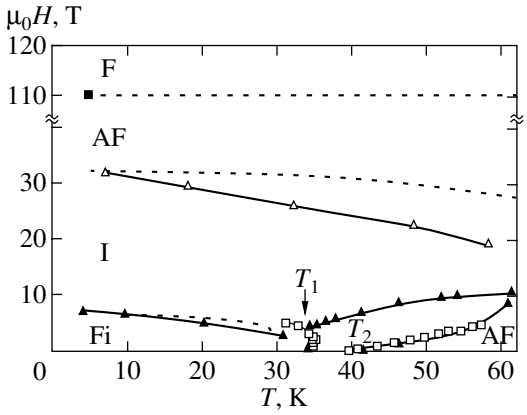

Fig. 6. Magnetic phase (H–T) diagram of $DyMn_2Ge_2$. Notation for phases is explained in the text. Dashed curves correspond to calculations. The remaining curves present experimental data obtained by us (dark square and light triangles), in [5] (dark triangles) and in [8] (light squares).

measurements exhibit a hysteresis. It can be seen that the calculated temperature dependence of the critical field $H_{Fi \to I}$ is in good agreement with the experimental curve obtained in [5] on a single crystal. The calculated field $H_{I \to AF}$ of the I ⟶ AF transition exhibits a weaker temperature dependence than the field obtained from measurements on free powders.

It was mentioned in the Introduction that the magnetic state of the intermetallic compound $DyMn_2Ge_2$ in zero field in the temperature interval between T_1 and T_2 has not been determined unambiguously in the available publications and the origin of the magnetic phase transitions in the temperature range above T_1 is not quite clear. It should be recalled that the results of the Mössbauer and neutron-diffraction studies carried out in [6] enabled the authors of this work to assume that three phases coexist in zero field in the temperature range between T_1 and T_2: ferrimagnetic (Fi), antiferromagnetic (AF), and intermediate (I) phases. This is possible if the intermediate phase in this temperature range is stable, while the other two phases are in the metastable state.

At the same time, Kobayashi *et al.* [7] concluded from the results of preliminary neutron diffraction studies that the magnetic structure in the temperature interval $T_1 < T < T_2$ is homogeneous: it has the form of an incommensurate phase with the wave vector $k = (0, 0, 0.65)$. Such a value of the wave vector indicates that this incommensurate phase is close to a commensurate phase with the unit cell tripling along the tetragonal c axis. In our opinion, this model is supported by the magnetization curves for $T > T_1$ from [5, 8], in which the hysteresis loops are small (if three different magnetic phases coexist in this temperature range, hysteresis effects must be manifested much more strongly). The version of the existence of an incommensurate phase suggests that the low-field magnetic phase transition observed in the antiferromagnetic phase at a temperature above T_2 is a transition to an incommensurate phase. This, however, is just a hypothesis, and we have to admit that the available experimental data are insufficient for describing theoretically the low-field magnetic behavior of $DyMn_2Ge_2$ above T_2.

Figure 7d demonstrates a fairly good agreement between the experimental magnetization curve [5] and the one calculated on the basis of the thermodynamic potential (5) for the AF phase in a field parallel to the tetragonal axis.

5. PHASE TRANSITIONS IN A FIELD PERPENDICULAR TO THE TETRAGONAL AXIS

The experimental H–T phase diagram for a field directed perpendicularly to the tetragonal axis has not been obtained, but the magnetization curves given in [5] for this direction at $T = 4.2$ and 20 K indicate the beginning of a first-order phase transition in a field (see Figs. 7a and 7b). At these temperatures, the Fi phase is initial. Our numerical calculations based on the thermodynamic potential (5) proved that, as expected, the susceptibility of the Fi phase for $\mathbf{H} \perp \mathbf{c}$ is extremely sensitive to the crystal field parameters. We had to modify the parameters from [6] within the error of their determination in this work and assume that (in cm^{-1})

$$B_2^0 = 200, \quad B_4^0 = -3.5,$$
$$B_6^0 = -50, \quad B_4^4 = -390,$$

in order to obtain simultaneously the susceptibility of the compound in the phase for $\mathbf{H} \perp \mathbf{c}$ and $T = 4.2$ K, coinciding with the experimental value from [5] (see Fig. 7a) without deteriorating the description of the regularities for the Mössbauer spectra given in [6]. In fields weaker than the transition field, the magnetic moment of dysprosium acquires a component along the magnetic field, while the magnetic moments of manganese remain parallel to each other and are slightly rotated towards the field direction since the exchange field exerted by dysprosium tends to preserve the antiparallel orientation of the moments of the dysprosium and manganese subsystems. In the critical field, a first-order phase transition to the triangular phase takes place, in which the magnetic moment of dysprosium is oriented along the field, while the moments of Mn form obtuse angles with it. The corresponding theoretical magnetization curves are presented in Figs. 7a and 7b. It can be seen that the calculated value of susceptibility in the Fi phase for 20 K coincides with the experimental value. We also succeeded in describing the decrease in the critical field upon heating (the phase-transition field at 20 K is weaker than at 4.2 K both in experiment and in theory). The calculated magnetization jump exceeds the experimentally observed value. Out attempts to increase the jump by slightly varying the parameters were futile. A change in the crystal field parameters,

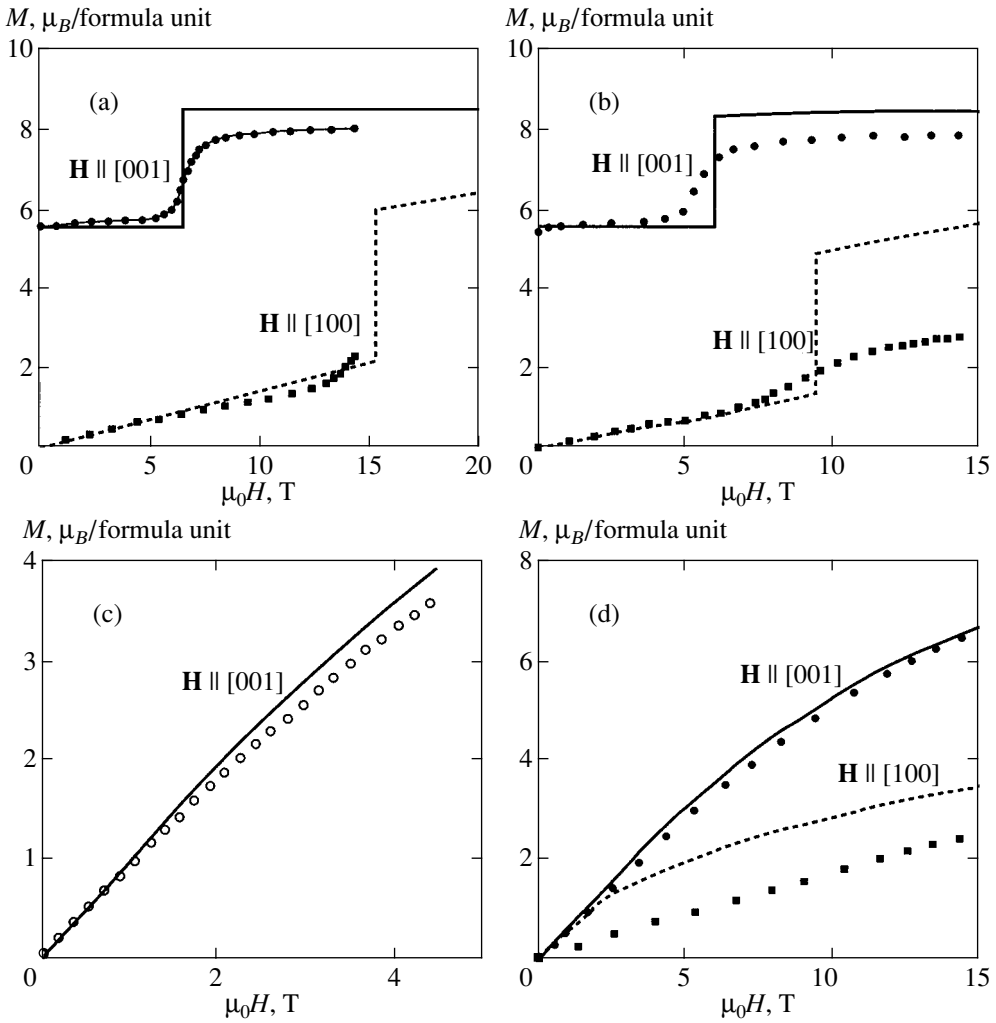

Fig. 7. Magnetization of $DyMn_2Ge_2$, calculated per formula unit as a function of the magnetic field applied along the indicated directions at various temperatures T, K: (a) 4.2; (b) 20; (c) 60, and (d) 77. The symbols correspond to the experimental results obtained in [5] (dark circles and squares) and [8] (light circles); the curves correspond to calculations.

e.g., a decrease in the parameter B_2^0, leads to an increase in the transverse susceptibility in the ferrimagnetic phase. The variations of the Dy–Mn and Mn–Mn exchange interaction parameters displace the phase transition fields for both directions of the field. The magnitude of the anisotropy constant for the manganese subsystem, which was taken, in particular, as 15.8 T μ_B/formula unit for $GdMn_2Ge_2$ [12], is insignificant since the anisotropy of the Mn subsystem is much smaller than the anisotropy of the Dy subsystem. Thus, the jump cannot be described in this model. A complication of the model would involve the introduction of new parameters whose determination makes the problem more ambiguous and requires additional experimental data for their determination.

It is interesting to note (see above) that the magnetic moment of dysprosium in a field perpendicular to the tetragonal axis can be reoriented relatively easily from the tetragonal axis to the direction of the field. For example, the angle formed by the magnetic moment of dysprosium with the tetragonal axis for H = 10 T amounts approximately to 7°. The tetragonal axis of the crystal must be the Ising axis for the moment of dysprosium at low temperatures since its ground state is characterized by the wave function $|15/2\rangle$. A comparatively easy reorientation of the magnetic moment of dysprosium from the Ising axis can be explained by the fact that the intersection of the lower energy levels of the Dy^{3+} ion (crossover) is observed for this direction of the field even in fields of about 1 T and the magnetic moment component perpendicular to the tetragonal axis starts growing at a higher rate.

Figures 7c and 7d show the magnetization curves for field directions in the basal plane in the temperature range in which the AF phase is initial. It can be seen that at T = 60 K (Fig. 7c), our theoretical curve is in good agreement with the experimental curve obtained in [8] during measurements on a thin square plate-like crys-

tal. We had to recalculate the experimental curve taking into account demagnetizing fields for the given shape of the single crystal, which was not done in [8]. At T = 77 K (Fig. 7d), the discrepancy between the calculated and experimental [5] magnetization curves is stronger. It cannot be ruled out that this is also associated with the necessity of including demagnetizing fields (there is no information in [5] about taking demagnetizing fields into account).

6. CONCLUSIONS

In this work, the layered tetragonal intermetallic compound $DyMn_2Ge_2$ is investigated experimentally and theoretically. The temperature of antiferromagnetic ordering of the manganese subsystem ($T_N \approx 440$ K) is determined from the measured temperature dependences of the lattice parameters a and c and the initial magnetic susceptibility. The existence of two low-temperature spontaneous first-order phase transitions ($T_1 \approx$ 33 K and $T_2 \approx 44$ K), which were established experimentally in [5, 6], is confirmed. The differential magnetic susceptibility at low temperatures is measured on free powders in strong magnetic fields (up to 50 T) and on power samples in ultrastrong magnetic fields (up to 150 T). Two first-order phase transitions are observed for the field orientation along the tetragonal axis: between 20 and 35 T (depending on temperature) and near 110 T.

Low-temperature magnetic properties of the layered intermetallic compound $DyMn_2Ge_2$ are explained in the model taking into account two magnetic subsystems of this intermetallic compound (the rare-earth and the manganese subsystems); the intrinsic exchange interaction of the manganese subsystem leads to antiferromagnetic ordering of this subsystem. It is shown that, in contrast to similar Gd-based intermetallic compounds studied earlier, the magnetic properties of $DyMn_2Ge_2$ can be interpreted correctly only by taking into account the exchange interaction between next-to-nearest Mn layers. This results in the possibility of the emergence, apart from antiferromagnetic and ferromagnetic types of magnetic ordering in the Mn subsystem, of a magnetic structure in which the mutual orientation of magnetic moments alternates as $\downarrow\uparrow\downarrow\downarrow\uparrow\downarrow\ldots$ in a field directed along the tetragonal axis. A comparison of the experimental values of induced phase transitions at helium temperature with the theoretically calculated values allowed us to determine the parameters describing both exchange interactions in the subsystems and the exchange interactions between the subsystems. Using these values of the exchange parameters as well as the parameters of the crystal field acting on the Dy^{3+} ion refined from the experimental data presented in [5, 6], we managed to describe semiquantitatively the magnetization curves at higher temperatures and to construct the H–T phase diagram.

It should be noted that some of the problems remain unsolved. For example, since the nature of magnetic ordering in the temperature range $T_1 < T < T_2$ has not be established completely, and the origin of field-induced low-field magnetic phase transitions at $T > T_1$ has not been clarified, the changes that must be introduced into the model for describing completely the magnetic phase diagram for $DyMn_2Ge_2$ are yet to be determined.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 99-02-17358) and by the program supporting science schools (grant no. 96-15-96429).

REFERENCES

1. K. S. V. Narasimhan, V. U. S. Rao, R. L. Bergner, *et al.*, J. Appl. Phys. **46**, 4957 (1975).
2. T. Shigeoka, J. Sci. Hiroshima Univ., Ser. A: Phys. Chem. **48**, 103 (1984).
3. A. Szutula and J. Leciejewicz, in *Handbook on the Physics and Chemistry of Rare Earths*, Ed. by K. A. Gschneidner, Jr. and L. Eyring (Elsevier, Amsterdam, 1989), Vol. 12, p. 133.
4. G. Venturini, B. Malaman, and E. Ressouche, J. Alloys Compd. **240**, 139 (1996).
5. H. Kobayashi, H. Onodera, Y. Yamaguchi, *et al.*, Phys. Rev. B **43**, 728 (1991).
6. G. Venturini, B. Malaman, K. Tomala, *et al.*, Phys. Rev. B **46**, 207 (1992).
7. H. Kobayashi, M. Ohashi, H. Onodera, *et al.*, J. Magn. Magn. Mater. **140**–**144**, 905 (1995).
8. H. Wada, Y. Tanabe, K. Hagiwara, and M. Shiga, J. Magn. Magn. Mater. **218**, 203 (2000).
9. A. S. Lagutin, J. Vanacken, N. Harrison, and F. Herlach, Rev. Sci. Instrum. **66**, 4267 (1995).
10. N. Puhlmann, H.-U. Mueller, M. Barczewski, *et al.*, J. Phys. D **32**, 2354 (1999).
11. Guo Guanghua, N. P. Kolmakova, R. Z. Levitin, *et al.*, in *Proceedings of MISM'99*, Ed. by A. Granovsky and N. Perov (Moscow State Univ., Moscow, 1999), Part 2, p. 133.
12. A. Yu. Sokolov, Guo Guanghua, S. A. Granovskiĭ, *et al.*, Zh. Éksp. Teor. Fiz. **116**, 1346 (1999) [JETP **89**, 723 (1999)].
13. Guo Guanghua, R. Z. Levitin, V. V. Snegirev, *et al.*, Zh. Éksp. Teor. Fiz. **117**, 1127 (2000) [JETP **90**, 979 (2000)].
14. C. Kittel, Phys. Rev. B **120**, 335 (1960).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 93, No. 4, 2001, pp. 805–808.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 921–925.

SOLIDS
Structure

Phase Stratification in the $Nd_{1-x}Ba_xCoO_{3-\delta}$ ($0.3 \le x \le 0.54$) System

D. D. Khalyavin[a, *], I. O. Troyanchuk[a], N. V. Kasper[a], and H. Szymczak[b]

[a]*Institute of Solid-State and Semiconductor Physics, Belarussian Academy of Sciences, ul. Brovki 17, Minsk, 220072 Belarus*

**e-mail: khalyav@ifttp.bas-net.by*

[b]*Institute of Physics, Polish Academy of Sciences, Warsaw, Poland*

Received April 26, 2001

Abstract—Slowly cooled $Nd_{1-x}Ba_xCoO_{3-\delta}$ samples were two-phase in the concentration interval $0.3 \le x \le 0.46$. One of the phases had *O*-orthorhombic lattice distortions (*Pbnm*) characteristic of ferromagnetic samples with $x \le 0.3$, and the other phase had tetragonal distortions (*P*4/*mmm*) characteristic of samples with $x \ge 0.46$. Tetragonal distortions were caused by ordering of Nd^{3+} and Ba^{2+} ions. Samples with ordered neodymium and barium ions ($Nd_{1-y}Ba_{1+y}Co_2O_{6-\gamma}$ at $-0.08 \le y \le 0.08$) experienced metal–dielectric and orientation magnetic phase transitions. © *2001 MAIK "Nauka/Interperiodica".*

Currently, substituted rare-earth metal orthocobaltites of the general formula $Ln_{1-x}A_x^{2+}Co_{1-x}^{3+}Co_x^{4+}O_3$ (Ln is a lanthanide, and A is a divalent alkaline-earth metal) attract much interest of researchers because their properties are in many respects similar to those of substituted orthomaganites $Ln_{1-x}A_x^{2+}Mn_{1-x}^{3+}Mn_x^{4+}O_3$, which exhibit giant magnetoresistance [1, 2]. In both classes of perovskites, an increase in the content of divalent ions causes a concentration transition to the ferromagnetic metallic state. Based on this observation, their magnetic properties were broadly interpreted in terms of double exchange [1, 2]. De Gennes [3] used double exchange theory to show that the concentration transition from the antiferromagnetic to the ferromagnetic state in $La_{1-x}A_xMnO_3$ at $0 < x < 0.2$ should occur via a noncollinear state. According to an alternative point of view on the mechanism of this transition, the system in the vicinity of the transition stratified into regions with different magnetic orderings [4, 5]. Experimental proof of magnetic phase stratification in manganites was obtained in [6]. The problem was studied theoretically in several works [7, 8]. Concentration transitions in substituted orthocobaltites have been studied much less thoroughly. The initial compounds, $LnCoO_3$, are an example of oxide systems with transitions between ion spin states. The splitting of the Co 3*d* levels by the ligand crystal field in these compounds is close in magnitude to the intraatomic exchange energy. For this reason, increasing temperature causes thermal excitation of trivalent cobalt from the low-spin $t_{2g}^6e_g^0$ state ($S = 0$) to the intermediate $t_{2g}^5e_g^1$ state with $S = 1$ or to the high-spin $t_{2g}^4e_g^2$ state ($S = 2$). As a consequence, the compounds are diamagnetic at low and paramagnetic at high temperatures [1]. It is shown in [9] that the replacement of neodymium by barium in $Nd_{1-x}Ba_xCoO_{3-\delta}$ causes a concentration transition from the diamagnetic to the ferromagnetic state and then from the ferromagnetic to the antiferromagnetic state. The special features of the transitions from the diamagnetic to the ferromagnetic state in this system are likely to be similar to those characteristic of the well studied $La_{1-x}Sr_xCoO_3$ system [1]. But information as to which of the possible mechanisms (noncollinear magnetic structure, spin glass-type state, or two-phase state) governs the transition from the ferromagnetic to the antiferromagnetic state is still lacking. The purpose of this work was to determine the mechanism of the concentration transition from the ferromagnetic to the antiferromagnetic state in the $Nd_{1-x}Ba_xCoO_{3-\delta}$ system.

All $Nd_{1-x}Ba_xCoO_{3-\delta}$ samples ($0.3 \le x \le 0.56$, concentration was varied in steps of $\Delta x = 0.02$) were synthesized by the usual ceramic procedure from metal oxides and carbonates of OSCh (special purity) grade taken in the required proportions. The mixtures were preliminarily calcined at 900°C. The syntheses were conducted at 1100–1200°C for 10 h. The samples were then slowly cooled (80 K/h) to ensure more complete absorption of oxygen by the lattice. To determine the effect of oxygen on the properties of samples, several samples were quenched from high temperatures. Electric conductivity was measured by the standard four-contact method. Magnetic measurements were performed on a commercial Foner magnetometer. X-ray diffraction was measured on a DRON-3 apparatus, CrK_α radiation.

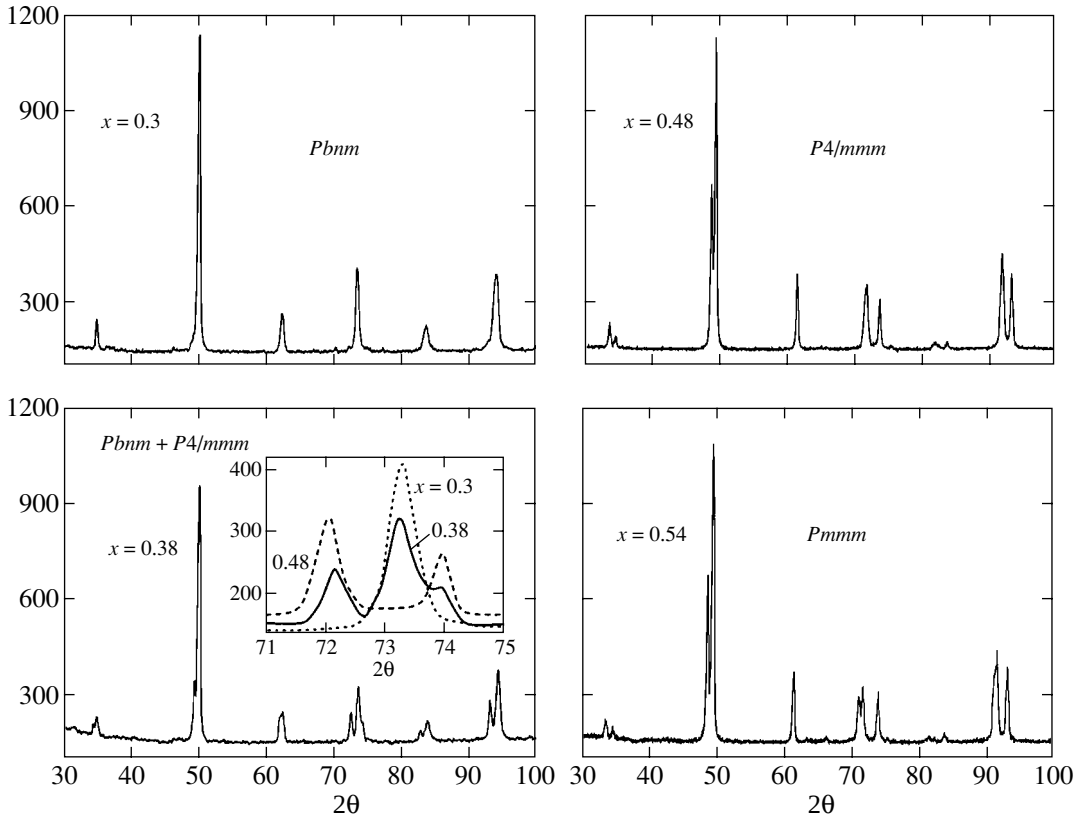

Fig. 1. X-ray powder patterns of some compositions of the $Nd_{1-x}Ba_xCoO_{3-\delta}$ series. Shown in the inset is the concentration evolution of the (200) multiplet in the cubic approximation.

According to the X-ray data, $Nd_{0.7}Ba_{0.3}CoO_{3-\delta}$, like initial $NdCoO_3$ [9, 10], has a perovskite structure with *O*-orthorhombic distortions, space group *Pbnm*. The X-ray powder pattern of $Nd_{0.68}Ba_{0.32}CoO_{3-\delta}$ contained not only reflections of the *O*-orthorhombic lattice but also lines of another phase. The intensity of these lines increased as the content of barium grew, whereas the intensity of the *O*-orthorhombic phase reflections simultaneously decreased. The system remained two-phase up to $x = 0.44$. The second phase, which appeared at $x > 0.3$, had a tetragonally distorted lattice, space group *P*4/*mmm*, characteristic of samples with barium contents $0.46 \le x < 0.5$. X-ray data therefore show that the $Nd_{1-x}Ba_xCoO_{3-\delta}$ system stratifies into regions with different crystal structure types in the concentration interval $0.3 < x < 0.46$ (Fig. 1). According to [9, 11], tetragonal distortions of the $Nd_{0.5}Ba_{0.5}CoO_{3-\delta}$ lattice are caused by spatial ordering of Nd^{3+} and Ba^{2+} ions similar to ordering of Y^{3+} and Ba^{2+} in $YBaFeCuO_5$ [12]. Ordering of ions allows $Nd_{0.5}Ba_{0.5}CoO_{3-\delta}$ to be treated as an individual compound whose chemical formula should be written as $NdBaCo_2O_{6-\gamma}$. According to the X-ray data, the compositions with ordered Nd and Ba ions, when slowly cooled in air (80 K/h), remain single-phase when the contents of Ba and Nd change from $Nd_{1.08}Ba_{0.92}Co_2O_{6-\gamma}$ to $Nd_{0.92}Ba_{1.08}Co_2O_{6-\gamma}$ ($Nd_{1-y}Ba_{1+y}Co_2O_{6-\gamma}$ with $-0.08 \le y \le 0.08$). In the $-0.08 \le y < 0$ interval, the compositions are characterized by tetragonal crystal lattice distortions (*P*4/*mmm*), and in the $0 \le y \le 0.08$ interval, distortions are orthorhombic (*Pmmm*). The degree of orthorhombic distortions decreases as barium content grows. Quenching from high temperatures substantially broadens the region of homogeneity of the ordered phase. We were able to obtain a single-phase $Nd_{1.34}Ba_{0.66}Co_2O_{6-\gamma}$ sample with tetragonal crystal lattice distortions by quenching this composition from $T = 1000°C$. This may be caused by an increase in oxygen deficiency in quenched samples, which stabilizes ordering of rare-earth metal ions and barium in perovskites [12].

Magnetization measurements showed that $Nd_{0.7}Ba_{0.3}CoO_{3-\delta}$ had spontaneous magnetization *M* below the critical temperature ~ 160 K. The magnetic moment of the cobalt sublattice (M_{Co}) was difficult to determine at low temperatures because of a negative *f*–*d* exchange value orienting the magnetic moment of the neodymium sublattice in the direction opposite to that of the cobalt sublattice. We were, however, able to estimate the upper bound for M_{Co} based of the field dependence of magnetization measured at a liquid helium temperature. The spontaneous magnetization of the sample with $x = 0.3$ at $T = 4.2$ K was $M = 0.5\mu_B$/formula unit. If we assume that, at this temperature, the neodymium sublattice is ordered with an $M_{Nd} = 1.4\mu_B$ magnetic

moment (the M_{Nd} value was obtained from the neutron diffraction data on $NdCoO_3$ [10] per Nd ion), then, for the Co sublattice, we obtain $M_{Co} = 1.4 \times 0.7 + 0.5 = 1.5\mu_B$/Co. The calculated M_{Co} value for the intermediate ($t_{2g}^5 e_g^1$, $S = 1$) spin state of Co^{3+} and the low-spin ($t_{2g}^5 e_g^1$, $S = 1/2$) state of Co^{4+} is $1.7\mu_B$/Co. It can therefore be suggested that some part of trivalent cobalt ions are in the low-spin state. On the other hand, the cobalt sublattice moment value lower than the calculated moment may be caused by an antiferromagnetic contribution of the Co^{3+}–O–Co^{3+} superexchange which competes with the Co^{3+}–O–Co^{4+} ferromagnetic exchange. The specific resistance of $Nd_{0.7}Ba_{0.3}CoO_{3-\delta}$ at $T = 77$ K equaled $\rho = 2 \times 10^{-3}\,\Omega$ cm and weakly increased as temperature rose, which is characteristic of metals (Fig. 2). An increase in the content of barium above $x = 0.3$ gradually decreased the spontaneous magnetic moment. Simultaneously, localization of charge carriers occurred, which was evidenced by the concentration transition from the metallic to the activated conductivity type (Fig. 2). For $Nd_{0.7}Ba_{0.3}CoO_3$, a magnetization peak was observed in measurements near the Curie point in low fields after zero field cooling (ZFC). The ZFC-magnetization curves of $Nd_{1-x}Ba_xCoO_{3-\delta}$ samples with $x > 0.3$ contained an anomaly in the low-temperature region ($T \sim 130$ K). This anomaly transformed into a well-defined additional peak as the concentration of barium increased. The relative intensities of both peaks changed depending on barium content (Fig. 3). It follows from the X-ray data that $Nd_{1-x}Ba_xCoO_{3-\delta}$ exists in the two-phase state in the concentration interval $0.3 < x < 0.46$. It is likely that the additional ZFC-magnetization peak at $x > 0.3$ is caused by the appearance of the tetragonal phase, in which barium and neodymium ions are ordered. The results obtained in studying the magnetic properties of the $Nd_{1-x}Ba_xCoO_{3-\delta}$ system can visually be represented in the form of the phase diagram shown in Fig. 4. The critical temperature for the $x = 0.2$ composition was taken from [9], for the other compositions, the critical temperatures were determined directly from the results of our experiments. The upper (solid) critical line, which separates the paramagnetic and magnetically ordered regions, corresponds to the onset of the transition from the paramagnetic to the ordered state. The lower (dotted) line corresponds to the position of the low-temperature maximum of the ZFC-magnetization curve (Fig. 3, inset). In the concentration range $0.3 < x < 0.4$, the characteristic temperatures of both phases vary insignificantly, whereas the intensity of the low-temperature magnetization peak gradually increases. Such behavior of the magnetic properties can be expected when the chemical composition of phases varies insignificantly, and changes occur as a result of an increase in the content of the tetragonal phase. In the $0.42 < x < 0.46$ interval, we observe inversion, that is, the temperature of magnetic ordering of the phase with ordered neody-

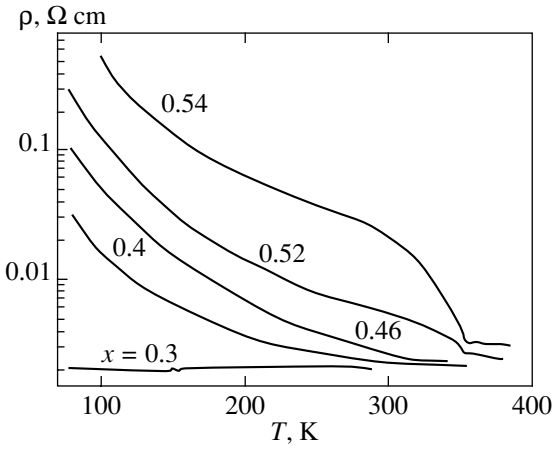

Fig. 2. Specific resistance of $Nd_{1-x}Ba_xCoO_{3-\delta}$ as a function of temperature.

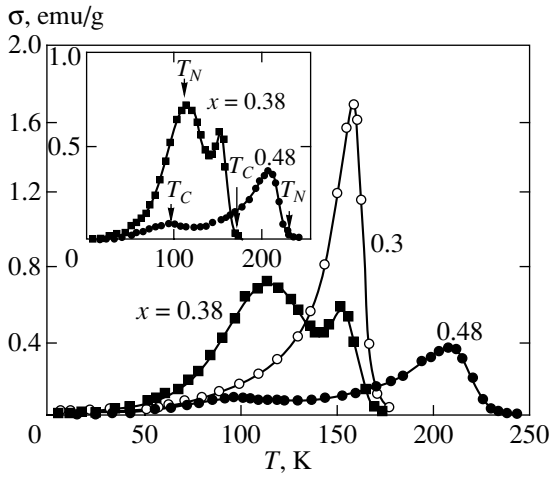

Fig. 3. Temperature dependences of ZFC-magnetization for samples of the $Nd_{1-x}Ba_xCoO_{3-\delta}$ series in an $H = 100$ Oe field.

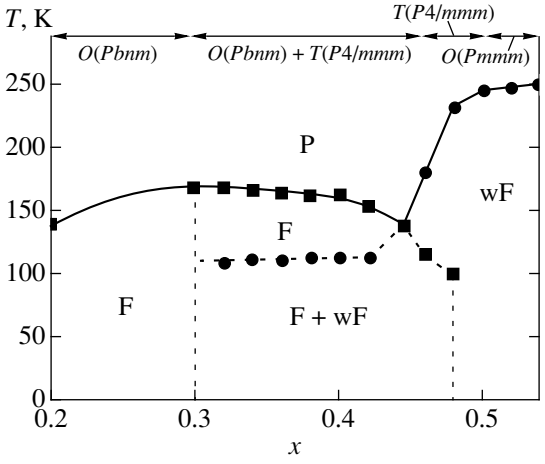

Fig. 4. Magnetic phase diagram of the $Nd_{1-x}Ba_xCoO_{3-\delta}$ system; F, ferromagnet; wF, weak ferromagnet; P, paramagnet; symbols ■ correspond to T_C and ●, to T_N.

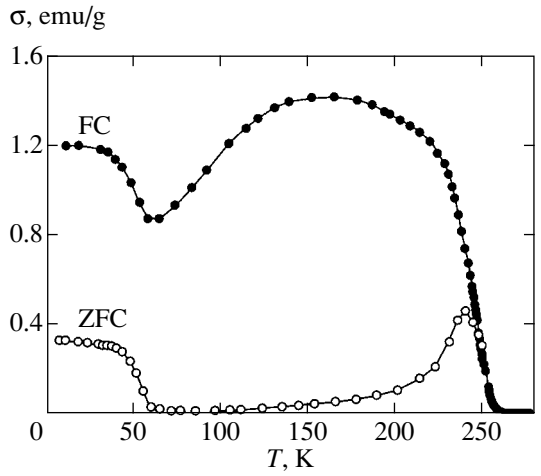

Fig. 5. Temperature dependences of FC- and ZFC-magnetizations in an $H = 100$ Oe field for $Nd_{0.92}Ba_{1.08}Co_2O_{6-\gamma}$.

mium and barium ions becomes higher than that of the second phase. The $Nd_{1-y}Ba_{1+y}Co_2O_{6-\delta}$ samples have spontaneous magnetizations not exceeding $0.15\mu_B$/Co. Currently, the nature of spontaneous magnetization of these compositions is subject to controversy. Phases with ordered barium and neodymium ions may well be weak ferromagnets. In the temperature range 30 K < T < 230 K, $Nd_{0.92}Ba_{1.08}Co_2O_{6-\delta}$ exhibits anomalous magnetization behavior in low magnetic fields (Fig. 5). This is likely to be related to orientation magnetic transitions, as follows from the observation that the magnetization measured after field cooling (FC) can be directed oppositely with respect to a low magnetic field in the temperature range 60–200 K. Orientation transitions have been thoroughly studied in rare-earth metal orthoferrites $RFeO_3$ [13]. In the high-temperature region, $Nd_{0.92}Ba_{1.08}Co_2O_{6-\gamma}$ has a well-defined metal–dielectric phase transition (Fig. 3), similar to that observed in $GdBaCo_2O_{5.5}$ [11]. It was shown in [14, 15] that orthorhombic crystal lattice distortions and the metal–dielectric transition in $RBaCo_2O_{6-\delta}$ (R is a rare-earth metal or yttrium ion) were closely related to the content of oxygen in these compounds and were characteristic of the $O_{5.5}$ phase, in which ordering of oxygen vacancies occurred. It is likely that deviations from the ratio between the cations in $NdBaCo_2O_{6-\gamma}$ toward increasing the content of barium causes an increase in oxygen nonstoichiometry and shifts γ closer to 5.5, similarly to what occurs in $RBaCo_2O_{6-\gamma}$ when the radius of the rare-earth metal ion decreases from Nd ($\gamma = 0.3$) to Tb ($\gamma = 0.5$) [14]. It follows from the data obtained in this work that the concentration transition from the ferromagnetic to the antiferromagnetic state in $Nd_{1-x}Ba_xCoO_{3-\delta}$ occurs through a mixed state, that is, the system stratifies into regions with different magnetic ordering types. The difference in magnetic ordering is caused by the different chemical compositions of these regions. Most likely, phase stratification occurs at temperatures about 600°C, when samples intensely interact with oxygen, which causes the decomposition of solid solutions.

ACKNOWLEDGMENTS

This work was financially supported by the Belarussian Foundation for Basic Research (project no. F99R-038) and Polish Committee for Science (grant 5 P03B 016 20).

REFERENCES

1. M. A. Senaris-Rodriguez and J. B. Goodenough, J. Solid State Chem. **118**, 323 (1995).
2. Y. Tokura and Y. Tomioka, J. Magn. Magn. Mater. **200**, 1 (1999).
3. P.-G. de Gennes, Phys. Rev. **118**, 141 (1960).
4. E. O. Wollan and W. C. Koehler, Phys. Rev. **100**, 545 (1955).
5. I. O. Troyanchuk, Zh. Éksp. Teor. Fiz. **102**, 251 (1992) [Sov. Phys. JETP **75**, 132 (1992)].
6. J. M. De Teresa, M. R. Ibarra, P. A. Algarabel, *et al.*, Nature **386**, 256 (1997).
7. S. Yunoki, A. Moreo, and E. Dagotto, Phys. Rev. Lett. **81** (25), 5612 (1998).
8. É. L. Nagaev, Fiz. Tverd. Tela (St. Petersburg) **40**, 2069 (1998) [Phys. Solid State **40**, 1873 (1998)].
9. I. O. Troyanchuk, N. V. Kasper, D. D. Khalyavin, *et al.*, J. Phys.: Condens. Matter **10**, 6381 (1998).
10. I. Plaza, E. Palacios, J. Bartolome, *et al.*, Physica B (Amsterdam) **234**, 632 (1997).
11. I. O. Troyanchuk, N. V. Kasper, D. D. Khalyavin, *et al.*, Phys. Rev. Lett. **80**, 3380 (1998).
12. L. Er-Rakho, C. Michel, Ph. Lacorre, and B. Raveau, J. Solid State Chem. **73**, 531 (1988).
13. K. P. Belov, A. K. Zvezdin, A. M. Kadomtseva, and R. Z. Levitin, *Orientational Transitions in Rare-Earth Magnetic* (Nauka, Moscow, 1979).
14. A. Maignan, C. Martin, D. Pelloquin, *et al.*, J. Solid State Chem. **142**, 247 (1999).
15. D. Akahoshi and Y. Ueda, J. Solid State Chem. **156**, 355 (2001).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 93, No. 4, 2001, pp. 809–814.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 926–932.

SOLIDS
Structure

Incommensurate Magnetic Structure in Copper Metaborate

G. A. Petrakovskii[a, *], M. A. Popov[b, **], B. Roessli[c], and B. Ouladdiaf[d]

[a] *Kirenskii Institute of Physics, Siberian Division, Russian Academy of Sciences, Akademgorodok, Krasnoyarsk, 660036 Russia*
**e-mail: gap@iph.krasn.ru*
[b] *Krasnoyarsk State University, Krasnoyarsk, 660041 Russia*
***e-mail: rsa@iph.krasn.ru*
[c] *Laboratory for Neutron Scattering, ETH Zurich and Paul Scherrer Institute, CH-5232 Villigen PSI, Switzerland*
[d] *Institut Laue–Langevin, 38042 Grenoble, Cedex 9, France*

Received April 24, 2001

Abstract—Based on the experimental data on copper metaborate single crystals obtained in X-ray and neutron diffraction studies and heat capacity, magnetic susceptibility, and muon spin relaxation measurements, a phenomenological theory of the incommensurate magnetic structure of this crystal was developed. Considering the space group of the crystal, $I\bar{4}2d$, Lifshits invariants were included into its thermodynamic potential. An analysis showed that magnetic structure formation at 10–20 K was dominated by the subsystem of copper spins in 4*b* unit cell sites. Below 10 K, the role played by the magnetic subsystem of copper spins in 8*d* unit cell sites in the formation of the magnetic structure of copper metaborate substantially increased. This caused a sharp increase in the wave vector of the incommensurate structure as temperature lowered. Numerical simulation of the temperature dependence of the wave vector of the helix and the heat capacity of the crystal gave a satisfactory description of the experimental data. This simulation was used to estimate the parameters of the phenomenological thermodynamic potential of the magnetic system of copper metaborate. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Modulated (incommensurate) magnetic structures are known to appear most often as a result of competition of exchange interactions [1]. The crystal structure then does not impose any limitations on the possibility of the formation of such structures. Less frequently, incommensurate structures are formed as a result of relativistic interactions. Dzyaloshinskii [2] was the first to mention such a possibility. The physical reason for the appearance of incommensurate structures is then the so-called Dzyaloshinskii–Moria antisymmetric exchange interaction. Formally, incommensurate structures of the relativistic origin can be described by introducing the Lifshits invariant [1], which linearly contains the first derivatives with respect to the two-component order parameter coordinates, into the thermodynamic potential. Note that an important limitation is then imposed on crystal symmetry. Namely, the Lifshits invariant can only be included into the thermodynamic potentials of crystals that have no center of symmetry. The magnetic state of a system with the Lifshits invariant is generally a lattice of magnetic solitons. In the simplest case, if magnetic crystallographic anisotropy effects are ignored, the distribution of magnetic moments obeys a simple sine law. The smallness of relativistic interactions implies the smallness of the wave vector of incommensurate magnetic structures. Close to the transition of a magnetic system into an incommensurate phase, strong diffuse neutron scattering should be observed [1]. The soliton character of magnetic ordering leads to a complex structure of magnetic satellites in the neutron diffraction pattern.

As modulated magnetic structures of the relativistic origin are rare, their detailed study, especially performed for high-quality single crystals, is of great interest. Copper metaborate CuB_2O_4 single crystals were for the first time synthesized and studied in [3–6]. Neutron diffraction data on this crystal were reported in [7]. In this work, we suggest a theoretical interpretation of the magnetic state of the spin system in copper metaborate at various temperatures; our interpretation is based on analyzing the whole set of available experimental data.

2. EXPERIMENTAL DATA

A procedure for growing high-quality large copper metaborate single crystals was described in [4]. According to X-ray and neutron diffraction studies [7], CuB_2O_4 forms tetragonal crystals, space group $I\bar{4}2d$ (D_{2d}^{12}) with lattice parameters a = 11.528 Å and c = 5.607 Å. The unit cell contains 12 formula units. Cop-

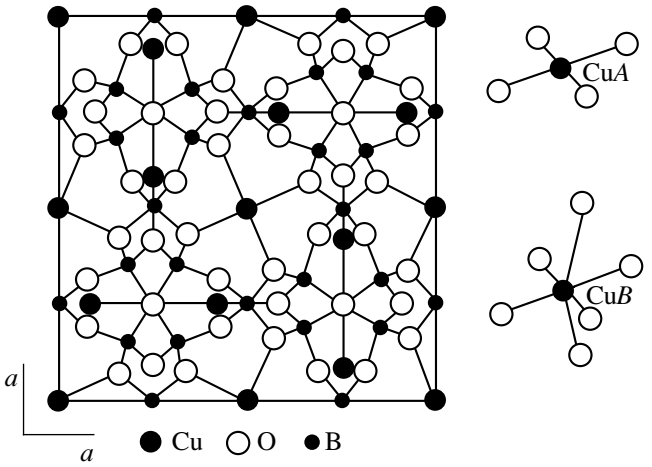

Fig. 1. Crystal structure of copper metaborate.

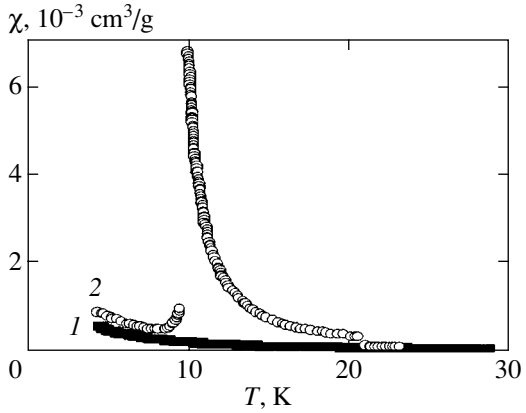

Fig. 2. Temperature dependence of magnetic susceptibility of single crystalline copper metaborate: (*1*) magnetic field is parallel to the tetragonal crystal axis and (*2*) magnetic field is parallel to the basal plane.

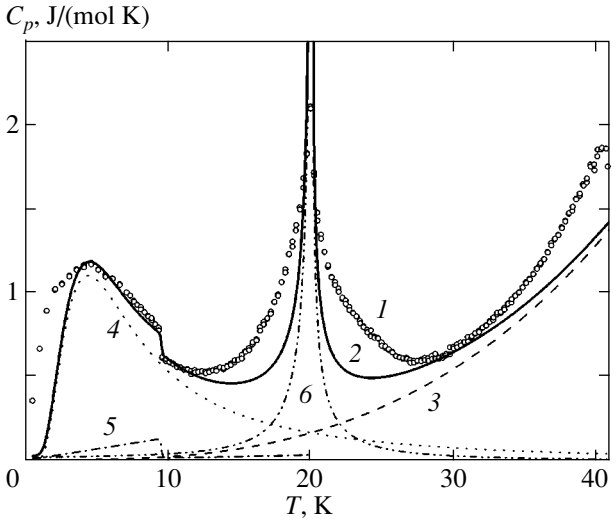

Fig. 3. Temperature dependence of the heat capacity of single crystalline copper metaborate: (*1*) experimental [3], (*2*) simulated, (*3*) Debye contribution, (*4*) Schottky anomaly, (*5*) Landau contribution, and (*6*) Ginzburg estimate.

per Cu^{2+} ions are situated in two nonequivalent positions, namely, Cu*A*, site 4*b*, point symmetry group S_4 (0, 0, 0.5) and Cu*B*, site 8*d*, point symmetry group C_2 (0.0815, 0.25, 0.125). Each Cu*A* ion occupies the center of a square formed by four oxygen ions. All Cu*B* ions are surrounded by six oxygen atoms in vertices of distorted octahedra (Fig. 1). The special high-resolution neutron diffractions study [7] showed that, up to 1.5 K, the crystal did not experience structural phase transitions of any kind. The magnetic reflections at 12 K corresponded to the Bragg commensurate phase positions. The observation of forbidden reflections of the (110) or (002) type at this temperature shows that the magnetic structure is antiferromagnetic. The magnetic and crystal chemical unit cells coincide, and the magnetic structure is described by the $\mathbf{q} = 0$ propagation vector.

Magnetic measurements on single crystals showed that sharp singularities appeared in the magnetic susceptibility curve at $T_A = 21$ K and $T_B = 10$ K. The temperature dependences of magnetic susceptibility obtained on a SQUID magnetometer for magnetic field orientations parallel and perpendicular to the tetragonal crystal axis are shown in Fig. 2 [6]. Attention is caught by the sharp anisotropy of susceptibility. For a magnetic field applied in the basal plane of the crystal, a susceptibility jump is observed at 21 K; susceptibility then rapidly increases as temperature decreases. At 10 K, susceptibility decreases in a jump by approximately one order of magnitude and then monotonically increases down to 4.2 K. For a magnetic field applied parallel to the tetragonal axis of the crystal, the temperature dependence of susceptibility is smooth in the whole temperature range. The Néel paramagnetic temperature and the effective copper ion moment determined from the high-temperature magnetic susceptibility portion equal $\Theta_N = -9.5$ K and $\mu_{\text{eff}} = 1.77\mu_B$, respectively.

The magnetic susceptibility anomalies described above are accompanied by singularities of the temperature dependence of heat capacity [3]. The results of heat capacity measurements in the temperature range 2–40 K are shown in Fig. 3. Two anomalies at the temperatures that coincide with those of magnetic susceptibility anomalies are quite manifest. In addition, the $C_p(T)$ curve contains an anomaly in the form of a broad maximum near 4 K.

Muon spin relaxation (μSR) data were reported in [3]. These data lend support to the conclusion of magnetic transformations at 21 and 10 K. More recent measurements at temperatures down to 1 K [8] revealed the occurrence of an additional magnetic transformation close to 1 K. This leads us to suggest that a new rearrangement occurs at this temperature in the spin subsystem of copper metaborate.

Neutron diffraction studies of copper metaborate were performed for a single crystal that contained the ^{11}B isotope to decrease absorption of neutrons [7]. The magnetic structure in the temperature range 10–21 K was found to be commensurate and antiferromagnetic.

This structure was described as a noncollinear arrangement of the Cu*A* and Cu*B* ion spins along the diagonals of the tetragonal plane of the crystal [7]. The magnetic moment of the Cu*A* ion was found to be about $1.3\mu_B$ at 12 K. Its component along the tetragonal *c* axis of the crystal was small, $\mu_z = 0.25\mu_B$, which corresponded to a 14° deviation from the *ab* plane. The Cu*B* ion spins lied in the *ab* plane and had a small magnetic moment of about $0.25\mu_B$ at $T = 12$ K. The magnetic moment of the Cu*B* ion rapidly increased as the temperature lowered below 10 K and equaled $0.7\mu_B$ at $T = 2$ K.

At temperatures below $T_B = 10$ K, two magnetic satellites situated symmetrically with respect to reciprocal lattice points of the commensurate phase appeared (Fig. 4). The magnetic structure of copper metaborate turned incommensurate along the tetragonal axis of the crystal and was described by a spin density wave with phase modulation [7]. The spin modulation period continuously increased from $\mathbf{q} \approx 0$ at 10 K to $\mathbf{q} = (0, 0, 0.15)$ at 1.8 K. At this temperature, spin structure modulation was characterized by a $c/0.15 \approx 40$ Å period. The temperature dependence of the wave vector of the incommensurate spin structure phase obeyed the power law

$$q(T) = A(T - T_B)^{0.48}. \quad (1)$$

As is shown in Fig. 5, Eq. (1) well described the temperature dependence of the incommensurate phase wave vector in the whole temperature range of measurements.

Wave vector **q** decreases virtually to zero as temperature increases, and the period of the incommensurate spin structure at T_B becomes large compared with the lattice parameter. In addition, strong diffuse neutron scattering is superimposed on the Bragg peaks for the $\mathbf{Q}_0$ neutron scattering vector along the [001] crystallographic direction. The intensity of diffuse scattering increases as temperature rises from 1.8 K and reaches a maximum close to T_B [7]. Diffuse scattering is observed even at the lowest temperature attained in experiments. This is sharply different from usual behavior of the spin subsystem of 3D-magnets with localized spins, for which critical fluctuations are limited by a narrow temperature region in the vicinity of the phase transitions. Adjustment of the magnetic structure at 2 K gives the best fit for a simple helix with a $0.7\mu_B$ for Cu*B*.

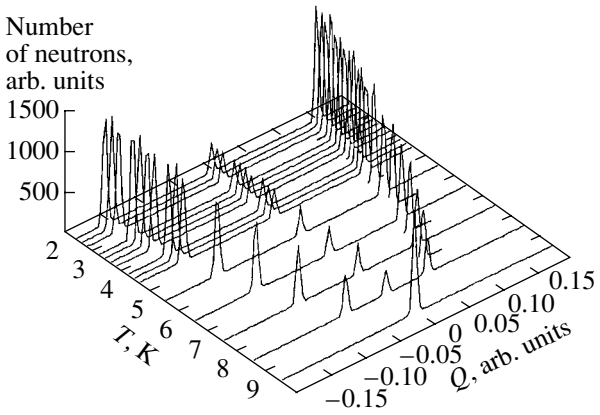

Fig. 4. Temperature dependence of propagation vector **q** = (3, 3, *Q*).

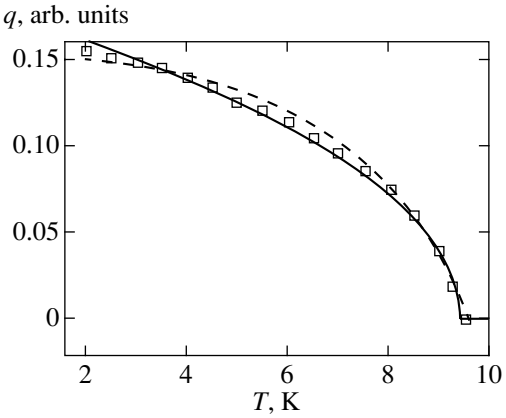

Fig. 5. Simulation of the temperature dependence of propagation vector *q* by Eqs. (1) (solid line) and (11) (dashed line).

3. DISCUSSION

The combined experimental data described above show that the spin subsystem of copper metaborate experiences rearrangements at 21 and 10 K. In addition, it may well be that one more rearrangement of the magnetic structure occurs near 1 K. In the commensurate phase, the magnetic structure is antiferromagnetic and possesses a spontaneous magnetic moment [6]. The magnetic and crystal chemical unit cells coincide, accordingly, the propagation vector **q** is zero. As the *I* symmetry operation of the lattice is also a magnetic transformation if $\mathbf{q} = 0$, the corresponding irreducible representations of the magnetic structure are the representations of the $\bar{4}2m$ point symmetry group. This group comprises eight symmetry elements and has five irreducible representations [9]. Four of these (Γ_1, Γ_2, Γ_3, and Γ_4) are one-dimensional, and one (Γ_5) is two-dimensional. The decomposition of representations gives

$$\Gamma_{4b} = \Gamma_3 + \Gamma_4 + 2\Gamma_5,$$

$$\Gamma_{8d} = \Gamma_1 + 2\Gamma_2 + \Gamma_3 + 2\Gamma_4 + 3\Gamma_5.$$

The Γ_3 and Γ_4 magnetic modes of site 4*b* correspond to collinear ferromagnetic and antiferromagnetic orderings along the *c* axis, respectively. The modes related to the Γ_5 representation describe a noncollinear magnetic structure. Similar magnetic modes for the 8*d* site can also be obtained from group theory. An analysis of the neutron diffraction pattern containing 25 purely mag-

netic peaks [7] shows that the magnetic structure of CuB_2O_4 copper metaborate can be described as a noncollinear arrangement of spins of both Cu*A* and Cu*B* along the diagonals of the tetragonal plane with out-of-plane displacements of Cu*A* magnetic moments. A symmetry analysis of the structure of copper metaborate shows that Cu*A* spins may be involved in Dzyaloshinskii–Moria interactions. Therefore, these interactions favor a noncollinear arrangement of spins, which is observed experimentally. In the incommensurate magnetic order region, helical ordering is observed [7].

It follows from the aforesaid that the whole spin system of copper metaborate comprises two subsystems, namely, subsystem *A* formed by Cu*A* ions, where the distribution of spins corresponds to a mixture of the Γ_4 and Γ_5 irreducible representations (axial antiferromagnetic and noncollinear planar configurations), and subsystem *B* formed by Cu*B* ion spins. The distribution of spins in this subsystem corresponds to the Γ_5 irreducible representation (a noncollinear planar configuration). It follows that, in writing the phenomenological thermodynamic potential of the spin system of copper metaborate, we must take into account two two-component order parameters corresponding to the Γ_5[Cu*A*] and Γ_5[Cu*B*] representations and one order parameter corresponding to the Γ_4[Cu*A*] representation. It is also important that crystal symmetry allows Lifshits invariants to be introduced for both subsystems.

Let the one-component order parameter be denoted by η, and two two-component order parameters, by (η_{A1}, η_{A2}) and (η_{B1}, η_{B2}) for subsystems *A* and *B*, respectively. The thermodynamic potential of the whole spin system can then be written as

$$\Phi = \int\left\{\frac{\alpha}{2}\eta^2 + \frac{\beta}{4}\eta^4 + \frac{\alpha_A}{2}\eta_A^2 + \frac{\beta_A}{4}\eta_A^4 + \frac{\gamma_A}{4}\eta_A^4\cos(4\varphi_A)\right.$$
$$+ \frac{\delta_A}{2}[(\nabla\eta_A)^2 + \eta_A^2(\nabla\varphi_A)^2] - \sigma_A\eta_A^2\varphi_A' + \frac{\alpha_B}{2}\eta_B^2 + \frac{\beta_B}{4}\eta_B^4$$
$$+ \frac{\gamma_B}{4}\eta_B^4\cos(4\varphi_B) \qquad (2)$$
$$+ \frac{\delta_B}{2}[(\nabla\eta_B)^2 + \eta_B^2(\nabla\varphi_A)^2] - \sigma_B\eta_B^2\varphi_B'$$
$$\left. + \kappa\eta_A\eta_B\cos(\varphi_A - \varphi_B) + \frac{\kappa_A}{2}\eta^2\eta_A^2 + \frac{\kappa_B}{2}\eta^2\eta_B^2\right\}dV,$$

where

$$\alpha_A = \alpha_{A0}(T - T_A), \quad \alpha_B = \alpha_{B0}(T - T_B),$$
$$\eta_A^2 = \eta_{A1}^2 + \eta_{A2}^2, \quad \eta_B^2 = \eta_{B1}^2 + \eta_{B2}^2,$$
$$\varphi_A = \arctan(\eta_{A2}/\eta_{A1}), \quad \varphi_B = \arctan(\eta_{B2}/\eta_{B1}),$$

$\alpha_{A0} > 0$, $\alpha_{B0} > 0$, $\beta_A > 0$, $\beta_B > 0$, $\delta_A > 0$, $\delta_B > 0$, ∇ is the nabla operator, and $f' \equiv df/dz$. The relation between subsystems *A* and *B* described by the invariant with coefficient κ in (2) results in the appearance of a helicoid in both subsystems at the same temperature.

The η one-component order parameter describes the axial component of Cu*A* spins and is of no importance for analyzing the incommensurate structure of the simple helix type with the helicoid axis along the tetragonal crystal axis. For this reason, thermodynamic potential (2) can be reduced to

$$\Phi = \int\left\{\frac{\alpha_A}{2}\eta_A^2 + \frac{\beta_A}{4}\eta_A^4 + \frac{\gamma_A}{4}\eta_A^4\cos(4\varphi_A)\right.$$
$$+ \frac{\delta_A}{2}[(\nabla\eta_A)^2 + \eta_A^2(\nabla\varphi_A)^2] - \sigma_A\eta_A^2\varphi_A'$$
$$+ \frac{\alpha_B}{2}\eta_B^2 + \frac{\beta_B}{4}\eta_B^4 + \frac{\gamma_B}{4}\eta_B^4\cos(4\varphi_B) \qquad (3)$$
$$+ \frac{\delta_B}{2}[(\nabla\eta_B)^2 + \eta_B^2(\nabla\varphi_A)^2] - \sigma_B\eta_B^2\varphi_B'$$
$$\left. + \kappa\eta_A\eta_B\cos(\varphi_A - \varphi_B)\right\}dV.$$

The Φ extremum conditions with respect to the order parameters corresponding to the equilibrium state of the system have the form

$$\frac{\delta\Phi}{\delta\eta_A} = \alpha_A\eta_A + \beta_A\eta_A^3 + \gamma_A\eta_A^3\cos(4\varphi_A) + \delta_A\eta_A(\nabla\varphi_A)^2 \qquad (4)$$
$$- 2\sigma_A\eta_A\varphi_A' - \delta_A\Delta\eta_A + \kappa\eta_B\cos(\varphi_A - \varphi_B) = 0,$$
$$\frac{\delta\Phi}{\delta\varphi_A} = -\gamma_A\eta_A^4\sin(4\varphi_A) - \delta_A\nabla(\eta_A^2\nabla\varphi_A) \qquad (5)$$
$$+ \sigma_A(\eta_A^2)' - \kappa\eta_A\eta_B\sin(\varphi_A - \varphi_B) = 0,$$
$$\frac{\delta\Phi}{\delta\eta_B} = \alpha_B\eta_B + \beta_B\eta_B^3 + \gamma_B\eta_B^3\cos(4\varphi_B) + \delta_B\eta_B(\nabla\varphi_B)^2 \qquad (6)$$
$$- 2\sigma_B\eta_B\varphi_B' - \delta_B\Delta\eta_B + \kappa\eta_A\cos(\varphi_A - \varphi_B) = 0,$$
$$\frac{\delta\Phi}{\delta\varphi_B} = -\gamma_B\eta_B^4\sin(4\varphi_B) - \delta_B\nabla(\eta_B^2\nabla\varphi_B) \qquad (7)$$
$$+ \sigma_B(\eta_B^2)' + \kappa\eta_A\eta_B\sin(\varphi_A - \varphi_B) = 0,$$

where Δ is the Laplase operator. As the perturbation of the homogeneous magnetic system state by the Lifshits invariant is one-dimensional, and deviations transverse to the *z* axis in the equilibrium state are excluded by positive δ_A and δ_B values, the nabla operator in these conditions should be replaced by single differentiation

with respect to z, and the Laplase operator, by double differentiation.

Finding the equilibrium state of the system from (4)–(7) is a complex problem. We will use the approximation of constant order parameter moduli, $\eta_A \neq \eta_A(z)$ and $\eta_B \neq \eta_B(z)$ [2]. As follows from (4) and (6), such an approximation is admissible if the Lifshits, anisotropy, and intersubsystem interaction invariants are small compared with the other invariants. As far as interaction between subsystems is concerned, it is only required that the dependence of the difference between helicoid phases in subsystems A and B on the z coordinate be negligibly small. In this approximation, thermodynamic potential (3) for the equilibrium state decomposes into two parts,

$$\Phi = \Phi_\eta + \Phi_\varphi,$$

$$\Phi_\eta = V\left\{\frac{\alpha_A}{2}\eta_A^2 + \frac{\beta_A}{4}\eta_A^4 + \frac{\alpha_B}{2}\eta_B^2 + \frac{\beta_B}{4}\eta_B^4 - |\kappa|\eta_A\eta_B\right\},$$

$$\Phi_\varphi = \int\left\{\frac{\gamma}{2}\cos(4\varphi) + \frac{\delta}{2}\varphi'^2 - \sigma\varphi'\right\}dV,$$

where $\varphi = \varphi_A = \varphi_B + \pi\theta(\kappa)$, $\theta(\kappa)$ is the Heaviside function,

$$\gamma = \gamma_A\eta_A^4 + \gamma_B\eta_B^4, \quad \delta = \delta_A\eta_A^2 + \delta_B\eta_B^2,$$

$$\sigma = \sigma_A\eta_A^2 + \sigma_B\eta_B^2.$$

Accordingly, the equilibrium conditions take the form

$$\begin{aligned}\alpha_A\eta_A + \beta_A\eta_A^3 - |\kappa|\eta_B &= 0,\\ \alpha_B\eta_B + \beta_B\eta_B^3 - |\kappa|\eta_A &= 0,\\ \delta\varphi'' + \gamma\sin(4\varphi) &= 0,\end{aligned} \tag{8}$$

where $f'' \equiv d^2f/dz^2$.

The first two equations of system (8) determine the temperature dependences of order parameter η_A and η_B moduli. Physically, the absence of phase $\varphi(z)$-dependent terms in these equations is justifiable, because η_A and η_B are almost fully determined by exchange interactions in the spin system of copper metaborate.

The third equation in (8) determines the dependence of the helicoid phase on the z coordinate along the tetragonal axis and on order parameter moduli. Its solution is the Jacobi amplitude function [10]

$$\varphi(z) = am(2q_0[z + \Delta z], k_1)/2, \tag{9}$$

where $q_0 = \sqrt{\gamma/\delta}/k_1$, and k_1 and Δz are the integration constants. The Δz constant corresponds to arbitrariness of selecting the origin along the tetragonal axis and is further set equal to zero. The k_1 constant is determined by the minimization of Φ_φ after substituting (9) into the equation for Φ_φ,

$$\begin{aligned}\Phi_\varphi &= \frac{V}{\lambda}\int_0^\lambda\left\{\frac{\gamma}{4}\cos(4\varphi) + \frac{\delta}{2}\varphi'^2 - \delta\varphi'\right\}dz\\ &= V\left\{\frac{4q_0\delta E(k_1) - 2\pi\sigma}{\lambda} + \frac{\gamma(1-2k_1^{-2})}{4}\right\},\end{aligned} \tag{10}$$

where

$$\lambda = \int_0^\lambda dz = \int_0^{2\pi}\frac{d\varphi}{|\varphi'|} = \frac{4K(k_1)}{q_0}, \tag{11}$$

is the helicoid period length, and

$$K(k_1) = \int_0^{\pi/2}\frac{du}{\sqrt{1-k_1^2\sin^2(u)}}$$

and

$$E(k_1) = \int_0^{\pi/2}\sqrt{1-k_1^2\sin^2(u)}\,du$$

are the total elliptical integrals of the first and second kind, respectively. The minimization of (10) with respect to k_1 gives

$$\frac{k_1}{E(k_1)} = \frac{2\sqrt{\gamma\delta}}{\pi\sigma}.$$

The corresponding Φ_φ value is

$$\Phi_\varphi = \frac{V\gamma(1-2k_1^{-2})}{4}.$$

Because of the absence of magnetic satellites in the spectra of inelastic neutron scattering on CuB_2O_4 in the temperature range 10–21 K, it can be suggested that $\sigma_A \ll \sigma_B$; therefore, σ_A can be ignored. The η_A order parameter at T_B is already not small, and anisotropy in subsystem A interferes with the appearance of a helicoid as a result of interaction between the subsystems. Therefore, the γ_A parameter is also assumed to be negligibly small in this work.

The relations obtained above allowed us to simulate the temperature dependences of η_A, η_B, and the $q = 2\pi/\lambda$ wave number and the temperature dependence of the specific heat capacity of the crystal, $C_p = -T\partial^2\Phi/\partial T^2$, with the use of the $C_p = -T\partial^2\Phi/\partial T^2$ relation for the following thermodynamic potential parameters (in kelvins):

$$\alpha_A = 1.6(T-20), \quad \beta_A = 2.7, \quad \gamma_A = 0,$$

$$\delta_A = 0.16, \quad \sigma_A = 0, \quad \kappa = 0.01,$$

$$\alpha_B = 4(T-9.6), \quad \beta_B = 2, \quad \gamma_B = 0.001,$$

$$\delta_B = 0.16, \quad \sigma_B = 0.1.$$

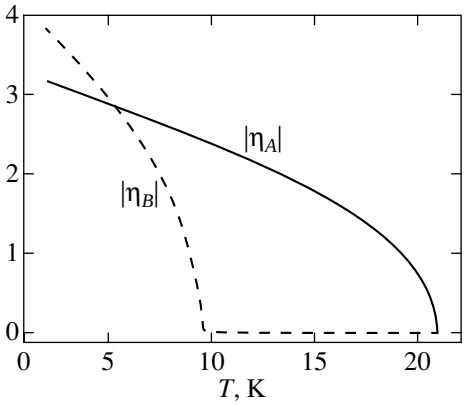

Fig. 6. Simulated temperature dependence of order parameter moduli.

Figure 6 shows that, in the temperature range 10–20 K, the η_B order parameter is small compared with η_A and rapidly increases at $T < 10$ K. A similar increase in the $q \equiv 2\pi/\lambda$ wave vector in Fig. 5 coincides with that observed experimentally (Figs. 4, 5) but qualitatively differs from the latter in that $q \neq 0$ already at temperatures below 20 K: in the temperature range of smallness of anisotropy invariants compared with the Lifshits invariants ($k_1 \ll 0$) we have

$$q \approx \sigma/\delta$$

and, at $\sigma_A = 0$, q variations are largely determined by η_B.

In addition to the jump at the transition point described by the Landau theory and the Ginzburg estimate of the contribution of order parameter thermal fluctuations, the contributions of acoustic phonons and Schottky-type anomalies were taken into account in heat capacity calculations. At 20 K, the fluctuation contribution characteristics of second-order phase transitions with a maximum at the transition point predominated. In the experimental curves (Fig. 3), the singularity at 9.6 K has the form of a step. This singularity is related to a rapid growth of the order parameter in the second spin subsystem. This parameter is already nonzero below 20 K because of bilinear interaction with the first subsystem. The field induced by this interaction suppresses thermal fluctuations in the second subsystem. The latter were therefore ignored in the calculations. Note that the step at 9.6 K is observed against the background of a broad maximum, which increases as temperature decreases. This maximum can be assigned to a Schottky-type anomaly that is not described by the phenomenological approach. The deviation of the experimental dependence from the calculation results at $T < 3.5$ K is caused by the approach to the supposed transition involving a low-temperature transformation of the magnetic structure near 1 K.

To summarize, we developed a phenomenological model of the magnetic subsystem of copper metaborate based on the experimental data on this compound. This model was applied to analyze the temperature dependence of the order parameter, the wave vector of magnetic structure modulation, and heat capacity. Further, we plan to study the properties of this crystal under strong magnetic field actions.

ACKNOWLEDGMENTS

The authors thank V.I. Zinenko for many valuable discussions. This work was financially supported by the Russian Foundation for Basic Research (project no. 01-02-17270) and Krasnoyarsk Regional Scientific Foundation (project no. 10F071C).

REFERENCES

1. Yu. A. Izyumov, *Neutron Diffraction on Long-Period Structures* (Énergoatomizdat, Moscow, 1987).
2. I. E. Dzyaloshinskii, Zh. Éksp. Teor. Fiz. **47**, 992 (1964) [Sov. Phys. JETP **20**, 665 (1964)].
3. G. Petrakovskii, D. Velikanov, A. Vorotinov, *et al.*, J. Magn. Magn. Mater. **205**, 105 (1999).
4. G. A. Petrakovskii, K. A. Sablina, D. A. Velikanov, *et al.*, Kristallografiya **45**, 926 (2000) [Crystallogr. Rep. **45**, 853 (2000)].
5. A. I. Pankrats, G. A. Petrakovskii, and N. V. Volkov, Fiz. Tverd. Tela (St. Petersburg) **42**, 93 (2000) [Phys. Solid State **42**, 96 (2000)].
6. G. A. Petrakovskii, A. D. Balaev, and A. M. Vorotynov, Fiz. Tverd. Tela (St. Petersburg) **42**, 313 (2000) [Phys. Solid State **42**, 321 (2000)].
7. B. Roessli, J. Schefer, G. Petrakovskii, *et al.*, Phys. Rev. Lett. **86**, 1885 (2001).
8. J. Schefer, B. Roessli, U. Staub, *et al.*, PSI Sci. Rep. **3**, 42 (1999).
9. O. V. Kovalev, *Irreducible and Induced Representations and Corepresentations of the Fedorov Groups* (Nauka, Moscow, 1986).
10. E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge Univ. Press, Cambridge, 1952; Fizmatgiz, Moscow, 1962), Part 1.

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 93, No. 4, 2001, pp. 815–823.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 933–944.

SOLIDS
Electronic Properties

Electrical Transport and Persistent Photoconductivity in Quantum Dot Layers in InAs/GaAs Structures

V. A. Kul'bachinskii[a, *], R. A. Lunin[a], V. G. Kytin[a], A. V. Golikov[a], A. V. Demin[a], V. A. Rogozin[a], B. N. Zvonkov[b], S. M. Nekorkin[b], and D. O. Filatov[b]

[a] *Moscow State University, Moscow, 119899 Russia*
**e-mail: kulb@mig.phys.msu.su*
[b] *Physicotechnical Research Institute, Nizhni Novgorod State University, Nizhni Novgorod, 603600 Russia*

Received February 14, 2001

Abstract—The conductivity of quantum dot layers is studied in InAs/GaAs structures in the temperature range from 300 to 0.05 K in the dark and using two types of illumination in magnetic fields up to 6 T. Depending on the initial concentration of current carriers, the conductivity of the structures varied from metallic (the Shubnikov–de Haas effect was observed) to hopping conductivity. At low temperatures, the temperature dependence of the resistance changed from the Mott dependence to the dependence described by the Shklovskii–Efros law for hopping conductivity in the presence of the Coulomb gap in the density of states. The conductivity of samples was studied upon their illumination at $\lambda = 791$ nm and $\lambda > 1120$ nm. All the samples exhibited a positive persistent photoconductivity at $T < 250$ K. The structures were also studied using photoluminescence and an atomic force microscope. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In recent years the quantum dot (QD) structures attract great scientific interest (see, for example, [1–4]). This is explained first of all by the possibility of studying fundamental physical laws in such structures, for example, localization and scattering of current carriers [5, 6] and optical properties [7, 8]. The growth processes and the energy spectrum are of interest in their own right [9–12]. The QD structures are also promising for practical applications in semiconductor lasers, transistors, and memory elements [13].

One of the promising methods for growing QD systems is the self-organization of InAs QD during the InAs deposition on the GaAs substrate, which is caused by the discrepancy between the periods of crystal lattices of InAs and GaAs. The deposition can be performed both by the method of molecular beam epitaxy and gaseous epitaxy from metalloorganic compounds (MOC hydride epitaxy) (see [3, 12, 14] and references therein). By varying the growth conditions, the size and density of QD in a layer can be changed. The properties of QD layers strongly depend on their size and distribution over the surface. The variation in the positions and sizes of QD can be reduced by growing them on the vicinal faces of a semiconductor [15].

At present, the optical properties of QD are studied quite thoroughly, whereas the transport of current carriers in QD structures has not been adequately investigated. In addition, the QD concentration in the structures was low in most studies in order to provide a sufficient overlap of the wave function of carriers localized at neighboring QD. In such structures, QD layers are not directly involved in photoconductivity, however, they can affect (as artificial scattering centers) the conductivity of neighboring two-dimensional layers [16, 17].

In this paper, we studied the conductivity in the plane of QD layers grown on the vicinal faces of the InAs/GaAs structures. The experiments were performed in the temperature range from 300 to 0.05 K in the dark and using two types of illumination in magnetic fields up to 6 T. At low concentrations of charge carriers, such structures exhibit the hopping conductivity, while at high concentrations of current carriers in QD layers, which are produced by doping, the mobility of current carriers is sufficient for observing oscillations of a magnetoresistance, the Shubnikov–de Haas effect.

2. SAMPLES

The structures were grown at the atmospheric pressure by MOC hydride epitaxy at the temperature 630°C on a vicinal semiinsulating GaAs substrate, i.e., the substrate was oriented at angles of 0.14° and 3° with respect to the (001) plane in the [110] direction. The structure is shown schematically in Fig. 1.

Samples of the main series with one QD layer (2704, 2706, 2711, 2932) consist (in the direction from the substrate to the surface) of the *i*-GaAs layer of thickness 0.45 μm, the Si δ layer, the undoped GaAs layer of thickness 18 nm (spacer separating the QD layer under study from the doping Si δ-layer), the InAs QD layer, another 18-nm-thick GaAs spacer and Si δ-layer, and the upper covering GaAs layer of thickness 0.45 μm. The Si δ-layers are required to provide the

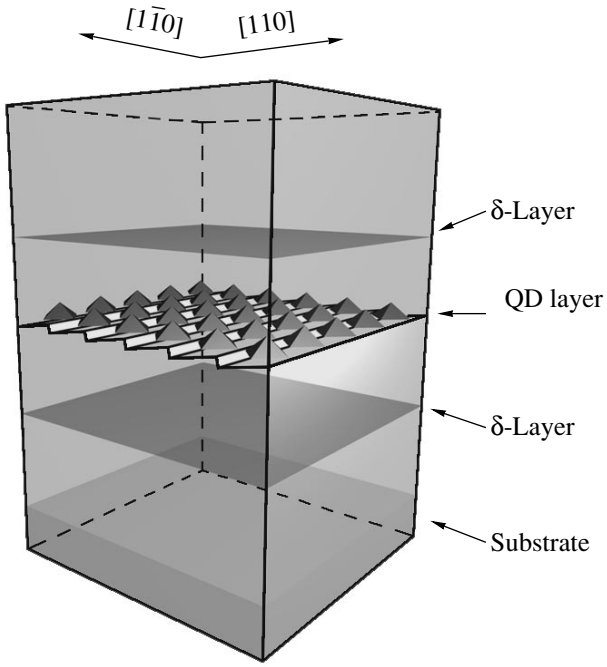

Fig. 1. Schematic structure of a sample with one QD layer.

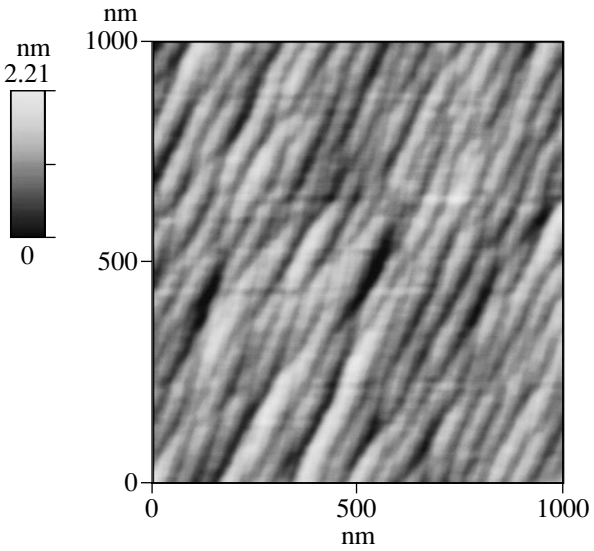

Fig. 2. Topography of a GaAs epitaxial layer observed with the help of an atomic force microscope.

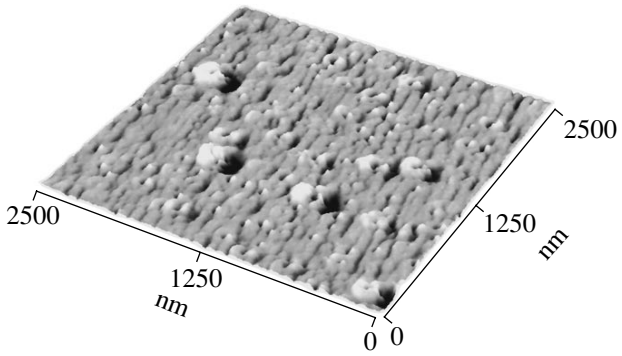

Fig. 3. An AFM image of the structure with InAs QD covered by a 10-nm thick GaAs layer.

electronic conductivity in QD layers. It is assumed that the δ-layers reduce the influence of the nonuniform distribution of the volume charge at the substrate interface and on the structure surface. The parameters of the structure were chosen so that the electronic levels of δ-layers lie substantially higher than the electronic levels of QD to exclude the shunting of QD layers by δ-layers. In sample 2713, according the photoluminescence data (discussed below), a QD layer was not formed because of an insufficient amount of deposited indium.

To make sure that the conductivity of the structures is realized through QD layers, we prepared test samples (2712 and 2715) without QD. Samples 2699 and 2700 were grown without δ-layers. Instead of this, sample 2699 was entirely uniformly doped except spacers near the QD layer, while sample 2700 contained broad (20 nm) uniformly doped regions. We also prepared samples 2933 and 2939 with two QD layers, separated by a GaAs layer of thickness 10 nm. Depending on the doping level, we could obtain various concentrations of electrons in the samples. Some parameters of the samples studied are presented in the table. We studied square samples with sides oriented along the [110] and [–110] directions (misorientation of the substrate results in the formation of steps oriented along the [–110] direction).

Along with samples with one or two QD layers, we prepared multilayer samples of the *p*- and *n*-types for studying the hopping conductivity (see table). The samples were grown by MOC hydride epitaxy at a temperature of 600–650°C and the atmospheric pressure on a semiinsulating GaAs substrate oriented at an angle of 3° with respect to the (001) plane in the [110] direction. Sample 1961 of the *p*-type consisted of ten GaAs periodic layers of thickness 0.1 μm, the carbon δ layer for the hole doping of QD layers, the GaAs spacer of thickness 5–6 nm, and ten InAs QD layers. The structures were covered from above by the 0.1-μm-thick GaAs layer. A great number of identical QD layers can be treated as parallel resistors. These samples were synthesized to reduce a high resistors of structures consisting of one layer for measuring resistances at low and ultralow temperatures.

The growth process and samples were studied with an atomic force microscope (AFM). Figure 2 shows the topography of the GaAs epitaxial layer. One can see a system of the growth macrosteps. Figure 3 shows the InAs QD structure covered by the GaAs layer 10 nm thick. The growth terraces and large islands are observed, the latter protruding over the cover layer and

Table

Sample	Sample type	Amount of InAs (ML)	E_{max}, eV	ΔE, meV	Illumination method	$N_H \times 10^{11}$, cm^{-2}	$N_{SH} \times 10^{11}$, cm^{-2}	μ_H, $cm^2/V\ s$
2712 *n*-type	Test sample without QD	0	–	–	In the dark	2.7	–	290
					Filter 1	4.2	–	1010
					Filter 2	4.3	–	1340
2715 *n*-type	Test sample without QD	0	–	–	In the dark	1.5	–	15
					Filter 1	3.9	–	300
					Filter 2	5.3	–	740
2699 *n*-type	Modulation doped One QD layer	4.5	1.352	42	In the dark	3.6	2.3	4900
			1.387	17	Filter 1	4	2.3	4900
					Filter 2	4.7	2.3	4900
2700 *n*-type	Modulation doped One QD layer	3.3	1.398	19	In the dark	2.5	2	6200
					Filter 1	3.4	2	6100
					Filter 2	3.7	2	6900
2713 *n*-type	Wetting InAs layer	1.4	1.471	13	In the dark	1.4	–	200
					Filter 1	3.3	–	1580
					Filter 2	4.5	–	1990
2704 *n*-type	One QD layer	3.3	1.393	20	In the dark	2.7	2.6	6300
					Filter 1	4.2	3.8	15 900
					Filter 2	4.3	4.5	32 000
2706 *n*-type	One QD layer	3.3	1.400	19	In the dark	–	–	–
					Filter 1	1.9	–	900
					Filter 2	3.2	–	1200
2711 *n*-type	One QD layer	3.3	1.397	26	In the dark	3	3.1	3400
					Filter 1	4.5	4.3	8200
					Filter 2	4.9	5.2	11 300
2932 *n*-type	One QD layer	3.3	1.37	23	In the dark	2.6	3.5	5700
					Filter 1	4.1	4.1	15 500
					Filter 2	5.3	4.5	22 700
2933 *n*-type	Two QD layers	3.3	1.362	26	In the dark	6	4.5	22 200
					Filter 1	6.2	4.5	23 100
					Filter 2	6.9	4.6	23 500
2939 *n*-type	Two QD layers	3.3	1.367	18	In the dark	2.8	2.7	8700
					Filter 1	4.5	3.5	10 500
					Filter 2	5.4	4.3	16 600
1961 *p*-type	Ten QD layers	2.5	1.36	22	In the dark	2.7	–	2700
1959 *n*-type	Ten QD layers	2.3	1.38	24	In the dark	1.9	2.6	10 500
1967 *n*-type	Twelve QD layers	2.1	1.41	26	In the dark	0.36	0.45	2400

Note: Parameters of samples: the amount of deposited InAs per layer (in QD monolayer (ML) units; positions E_{max} and half-widths ΔE of luminescence bands measured at 77 K; electron concentrations N_H (per layer) and N_{SH} measured from the Hall and Shubnikov–de Haas effects, respectively; Hall mobility μ_H measured in the dark and upon illumination through silicon filter no. 1 ($\lambda \geq 1120$ nm) and through filter no. 2 ($\lambda = 791 \pm 8$ nm) at 4.2 K.

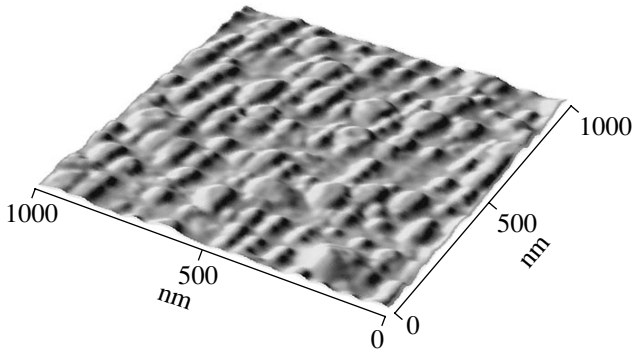

Fig. 4. An AFM image of the structure surface after selective etching for removing a cover layer.

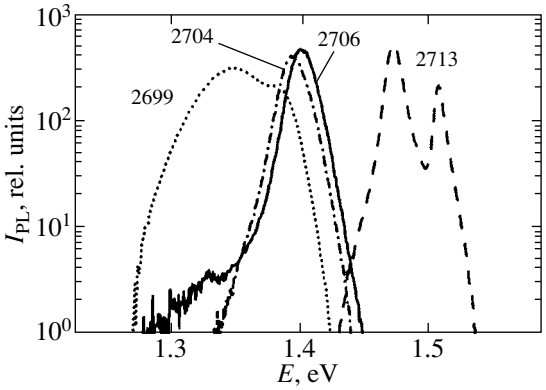

Fig. 5. Photoluminescence spectra of samples 2699, 2704, 2706, and 2713. The 1.52-eV peak corresponds to a bulk GaAs.

having diameters of about 250 nm, heights up to 100 nm, and the surface density equal to 10^8 cm^{-2}. The characteristic feature of the relief is a lot of craters in the cover layer ($N_s \sim 10^9$ cm^{-2}) and at the apexes of almost all islands. These craters suggest that the dissolving of relatively large dislocated island clusters takes place upon deposition of the cover layer, small pseudomorphous clusters representing QD being preserved, although it is possible that their shape and size somewhat change. To reveal QD, we removed the cover layer of the prepared samples by selective etching. Figure 4 shows the surface structure after this procedure. The QD are located along the terraces of the vicinal surface used for their growth.

3. RESULTS AND DISCUSSION

3.1. Photoluminescence

Photoluminescence spectra were recorded at 77 K by exciting samples with 200 mW/cm^2 He–Ne laser. The spectra of some samples are shown in Fig. 5. The table presents the energies of maxima of the spectra (spectra of some samples had two maxima) and the half-width of the spectral bands. According to the results obtained in [18], these data allows one to estimate approximately the QD size. In our case, such an estimate gives the diameter of a QD equal to 9–13 nm. Analysis of the galvanomagnetic properties of the structures studied suggests that in our case the wave functions of current carriers localized at neighboring QD (QD are densely located) strongly overlap with each other. As the distance between QD decreases, the luminescence bands shift to the red [19]. Thus, the estimate of the QD size from data [18] yields the upper bound. The direct study of QD structures with the help of an atomic force microscope gives the QD size equal to 5–6 nm.

The luminescence spectrum of sample 2713 (Fig. 5) corresponds to one InAs monolayer (wetting layer), i.e., in this sample the QD have not been formed because of a small amount of indium used for the sample growth. In samples 2706, 2707, and 2699, QD have already been formed. The red shift of the luminescence band of sample 2699 is caused by a great amount of deposited indium arsenide (see table) and, hence, by larger QD. The luminescence spectrum of this sample can be approximated by two bands, the long-wavelength band being broader, which suggests that the size distribution of QD is broad.

3.2. Anisotropy of Resistance and the Hopping Conductivity

We studied the temperature dependences of the resistance of all the samples. As an example, Fig. 6 shows the temperature dependences of the specific resistance for samples 2933 and 2711. One can see that the conductivity has anisotropy, the conductivity of all the samples along steps (the [–110] direction) being greater than that across the steps (inset in Fig. 6a). We explain this anisotropy by the appearance of chains formed by QD, which are oriented along the steps (along the [–110] direction) (see also [20]). Such anisotropy of the conductivity was observed on vicinal surfaces of GaAs structures δ-doped with tin [21, 22]. The predominant distribution of tin along the steps observed in papers [21, 22] resulted in the temperature-dependent anisotropy of the conductivity, and the conductivity along the steps was also greater.

At sufficiently low temperature and low concentration, current carriers are localized on the inhomogeneities of the QD layer and the conductivity occurs through hops of the current carriers from one localization center to another. The temperature dependence of the resistance in the case of hopping conductivity is determined by the law [23]

$$\rho = \rho_0 \exp\{(T_0/T)^{1/\nu}\} \quad (1)$$

(the Mott law for the variable range hopping conductivity) in the two-dimensional case at the constant density of states near the Fermi level $\nu = 3$. As the temperature

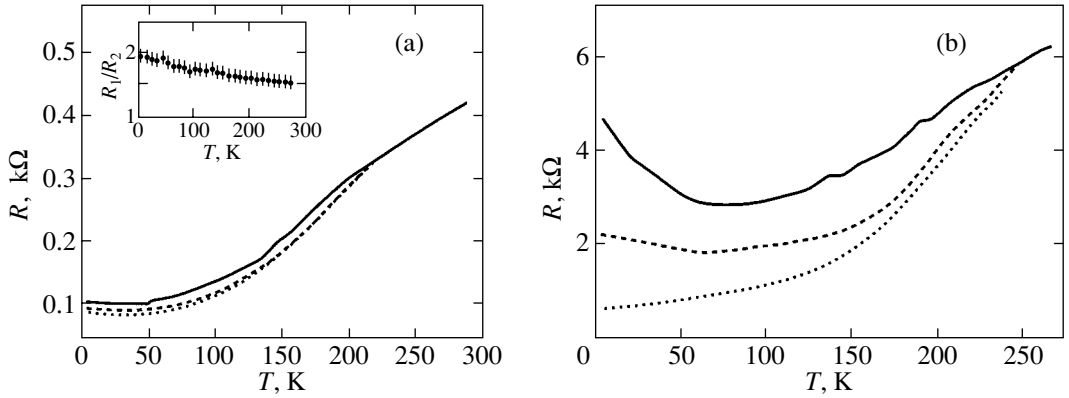

Fig. 6. Temperature dependences of the sheet resistance for samples (a) 2933 and (b) 2711. The solid curve was obtained in the dark, the dashed and dotted curves correspond to illumination through filters no. 1 ($\lambda \geq 1120$ nm) and no. 2 ($\lambda = 791 \pm 8$ nm), respectively. The inset shows the temperature dependence of anisotropy of the resistance of sample 2933.

decreases, the Coulomb interaction between current carriers becomes substantial, resulting in the formation of an energy gap in the spectrum of the density of states near the Fermi level [23]. When the maximum energy of a hop becomes equal to the energy gap, the dependence with $\nu = 3$ changes to the dependence with $\nu = 2$ (the Shklovskii–Efros law for the hopping conductivity in the presence of the Coulomb gap in the density of states), which was experimentally observed in δ-doped $GaAs/Al_xGa_{1-x}As$ heterostructures [24].

The resistance of samples with a sufficiently high mobility of current carriers ($\mu > 1000$ cm^2/V s at 4.2 K) decreased with decreasing temperature, passed through a minimum in the region between 30 and 100 K and again increased. The resistance of samples with a low mobility of carriers monotonically increased with decreasing temperature, and its temperature dependence was well described by expression (1) at sufficiently low temperatures (hopping conductivity).

We determined the parameter ν by plotting the temperature dependences of the resistance of samples in special coordinates $W(t)$, where $W = -\partial \ln\rho/\partial \ln T = (1/\nu)(T_0/T)^{1/\nu}$. In these coordinates, the parameter ν is determined from the derivative of the function $W(T)$ (the method is described in paper [24]). Figure 7 presents the dependence $W(T)$ for multilayer sample 1961 of the *p*-type ($p = 2.7 \times 10^{11}$ cm^{-2}) in logarithmic coordinates. As the temperatures decreases, dependence (1) with $\nu = 3$ transforms to the dependence with $\nu = 2$ at $T = 3.2$ K, which, according to the theory [23], occurs in the presence of a gap in the density of states at the Fermi level at a sufficiently low temperature. In other words, the temperature dependence of the resistance obeying the Mott law for the hopping conductivity in the two-dimensional case

$$\rho(T) = \rho_0 \exp(T_M/T)^{1/3},$$

with $T_M = 15$ K changes to the temperature dependence

$$\rho(T) = \rho_0 \exp(T_{ES}/T)^{1/2},$$

with $T_{ES} = 3.8$ K corresponding to the Shklovskii–Efros law. The resistance of multilayer samples of the *n*-type increased with decreasing temperature down to 0.05 K, as shown in Fig. 8, however, it could be approximated by expression (1) neither with $\nu = 3$ nor 2, which is, probably, explained by a lower effective mass of electrons and their substantially greater mobility compared to holes.

3.3. Persistent Photoconductivity

We studied the conductivity of samples in the dark and upon illumination at different wavelengths in the temperature range from 4.2 to 300 K. The samples were illuminated by light that passed through filter no. 1 ($\lambda \geq 1120$ nm, Si plate) or filter no. 2 ($\lambda = 791 \pm 8$ nm, interference filter). To determine the concentration and mobility of current carriers, we measured the Hall effect in a magnetic field up to 6 T at 4.2 K.

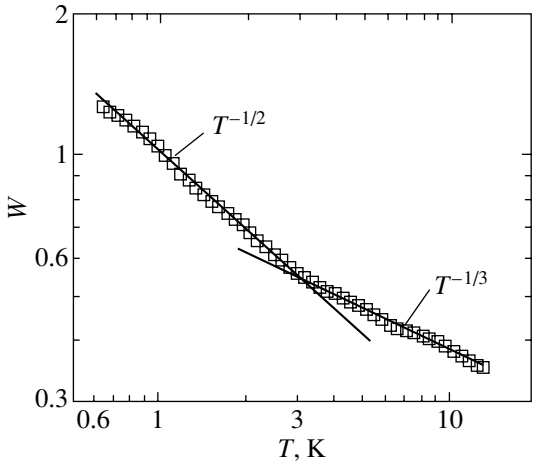

Fig. 7. Temperature dependence of logarithmic derivative $W = -\partial \ln\rho/\partial \ln T$ for sample 1961.

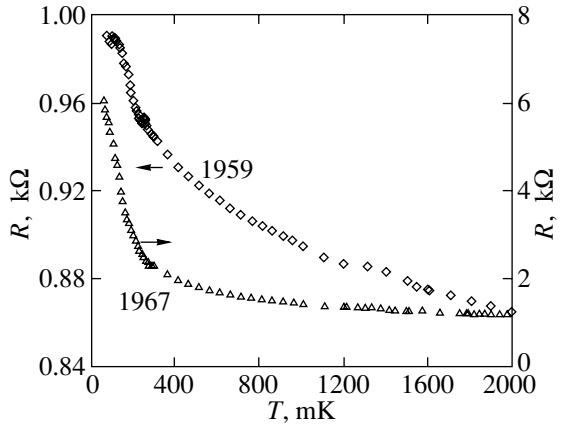

Fig. 8. Temperature dependences of the resistance of samples 1959 and 1967 of the *n*-type in the region of ultralow temperatures.

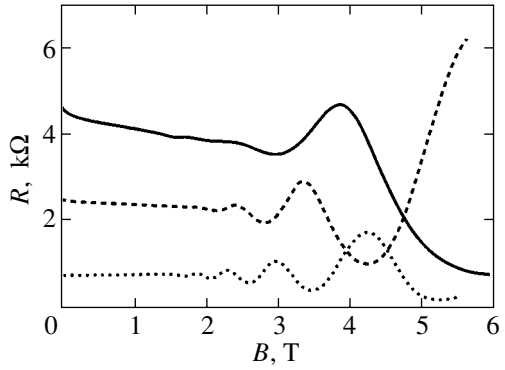

Fig. 9. Oscillations of the magnetoresistance of sample 2711 at 4.2 K. The solid curve corresponds to measurements in the dark. The dashed and dotted curves were obtained upon illumination through filters no. 1 ($\lambda \geq 1120$ nm) and no. 2 ($\lambda = 791 \pm 8$ nm), respectively.

An important remark should be made concerning the possibility of calculation of the concentration n of current carriers in our samples at 4.2 K from the Hall effect. We used two methods for measuring the concentration of current carriers that were based on the Shubnikov–de Haas effect (see below) or the Hall effect. In the case of the Hall effect, the classical formula $n = 1/R_H e$ (where R_H is the Hall coefficient) was used, which is valid for completely delocalized current carriers. However, at low temperatures some of the samples exhibited the hopping conductivity. The theory of the Hall effect in the case of hopping conductivity is much more complicated than the classical theory [25, 26]. Nevertheless, the classical formula is also valid in this case in certain regions of the phase diagram [26], which is confirmed experimentally (for example, in paper [27]), although in some cases this formula gives only approximate values of the concentration. Here, we use the concentration of current carriers calculated from the Hall effect for samples exhibiting hopping conductivity only as an estimate in a comparative analysis of the samples.

At low temperatures, we observed the positive persistent photoconductivity in all the samples. The typical dependences of the resistance on temperature and illumination are shown in Fig. 6. The sample was illuminated to achieve the saturation of its resistance, then the illumination was switched off and the temperature dependence of the resistance was measured with the rate 5 K/min. The temperature dependence of the difference between resistances measured in the dark and after illumination did not exhibit a threshold in all the samples studied. However, at $T > 250$ K, this difference became less than the error of measurements.

The conductivity of all the samples increased after their illumination. The Hall concentrations and mobilities generally increased after illumination, however, some of the samples revealed no changes (see table). The conductivity, Hall concentration and mobility increased greater after illumination through filter no. 2 compared to filter no. 1. In addition, upon illumination through filter no. 2, the conductivity saturated substantially faster.

Samples with a relatively high Hall mobility exhibited at low temperatures oscillations of a magnetoresistance (the Shubnikov–de Haas effect) from two-dimensional electrons in QD layers. The two-dimensional properties of electrons were verified experimentally by deviating the magnetic filed direction from the normal to the structure surface. Typical oscillations observed in some samples are shown in Figs. 9 and 10. The oscillation spectrum of each sample has a single frequency in the range of magnetic fields studied, which allows us to calculate easily the electron concentration and its dependence on illumination. The electron concentrations determined from the Shubnikov–de Haas effect are presented in the table.

The type of dependence of the oscillation frequency on illumination is the same for all the samples (except test samples 2699, 2700, 2712, and 2715). The oscillation frequency of the magnetoresistance increases upon illumination (Fig. 9), this increase being greater upon illumination through filter no. 2 compared to filter no. 1. In addition, upon illumination, oscillations begin at lower magnetic fields than in the dark, which suggests that the mobility of electrons increases. The increase in the Hall mobility with increasing concentration of current carriers in the QD layer can be explained by a stronger screening of the random scattering potential by electrons and by the increase in the Fermi velocity of electrons.

It is known that current carriers localized in Si δ-layers have a low mobility ($\mu < 2000$ cm^2/V s) [28]. In test samples 2712 and 2715 (samples with δ-layers, but without QD layers) with a very low mobility of current carriers (see table), no Shubnikov–de Haas oscillations were observed in magnetic fields up to 6 T. In the case of a relatively low doping level, all the electrons from

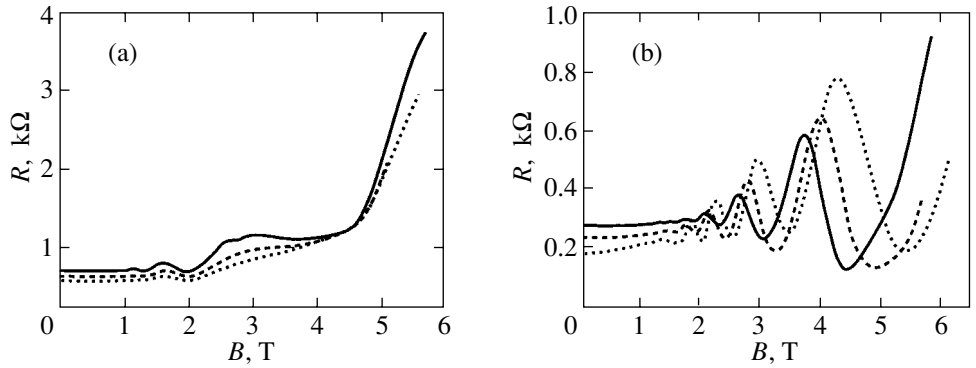

Fig. 10. Oscillations of the magnetoresistance of samples (a) 2699 and (b) 2933 at 4.2 K. The solid curve corresponds to measurements in the dark. The dashed and dotted curves were obtained upon illumination through filters no. 1 ($\lambda \geq 1120$ nm) and no. 2 ($\lambda = 791 \pm 8$ nm), respectively.

δ-layers transfer to QD layers. When the doping level increases, a fraction of electrons can remain in δ-layers, however, they do not contribute to the Shubnikov–de Haas oscillations. In other words, only two-dimensional electrons located in the QD layer will be involved in the Shubnikov–de Haas effect. The mobility of current carriers observed in QD structures substantially exceeds the mobility typical of δ-layers. Therefore, the presence of δ-layers virtually does not affect the transport properties of QD samples.

When a sufficiently high electron concentration is produced (with respect to the capacity of QD layers), the QD layer can be completely filled, and a fraction of current carriers will remain in δ-layers. The illumination will no longer change the concentration of carriers in the QD layer, but it can add electrons to δ-layers. Because of the low mobility of current carriers in δ-layers, the total conductivity is mainly determined by electrons located in QD layers. Therefore, the Hall mobility increases only slightly, whereas the Hall concentration increases. The frequency of Shubnikov–de Haas oscillations will not change noticeably because the electron concentration in the QD layer does not change. Most likely this situation is realized in samples 2699 and 2700 (Fig. 10a) and partially in sample 2933 (Fig. 10b), whereas in weakly doped samples the oscillation frequency increases upon illumination (Fig. 9). Sample 2933 was heavily doped, so that its QD layers were virtually filled even in the dark. Samples 2699 and 2700 contained one QD layer, but instead of δ-doping they were uniformly doped, resulting in a higher electron concentration in the QD layer.

Figure 11a shows schematically the band diagram of the samples in the absence of illumination. The Fermi level in a substrate is determined by the chromium level, which is located higher by 0.89 eV than the top of the valence band in GaAs [29]. Near the semiconductor surface, the Fermi level is determined by crystal defects. Because surface states capture electrons, the bands are bent near the surface, as shown in Fig. 11a.

We used two types of illumination in photoconductivity experiments. The light transmitted by filter no. 2 ($\lambda = 791 \pm 8$ nm) has a sufficient energy (the photon energy is $\varepsilon = 1.6$ eV) to excite an electron from the valence band to the conductivity band (the energy gap of GaAs at 4.2 K is 1.52 eV). The energy of photons transmitted by filter no. 1 ($\lambda \geq 1120$ nm, $\varepsilon_\gamma < 1.1$ eV) is not sufficient for exciting an electron from the valence band even to the electronic level of the size quantization (the luminescence band of QD lies in the region between 1.35 and 1.47 eV), but it is sufficient for exciting an electron from the chromium level in the substrate.

Upon illumination of a sample by light through filter no. 2, electron–hole pairs are formed over the entire sample thickness, from the substrate to the sample surface. Taking into account the energy band relief, the electrons will pass to QD layers and δ-layers, and then the electrons will tunnel from δ-layers to the QD layer (the distance between the δ-layers of the dopant and

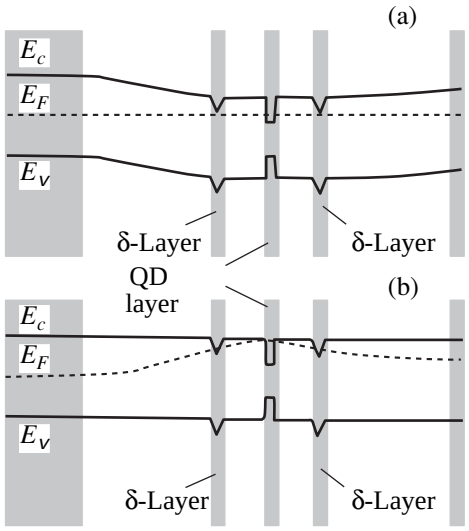

Fig. 11. Scheme of energy bands (a) before illumination and (b) after the resistance saturation.

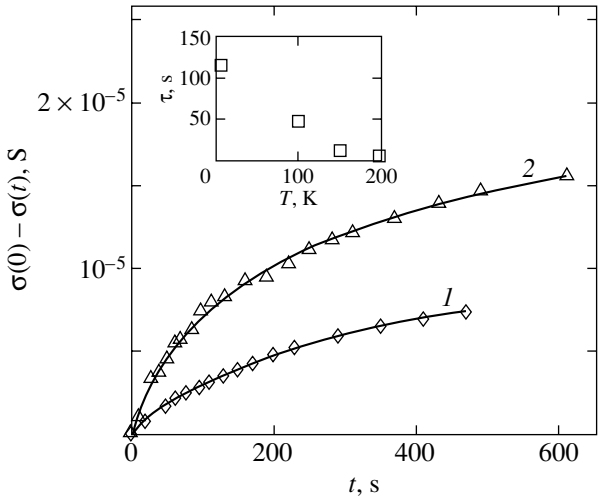

Fig. 12. Relaxation of the conductivity $\sigma(0) - \sigma(t)$ of sample 2713 measured in the dark at 4.2 K after illumination through filters (*1*) no. 1 ($\lambda \geq 1120$ nm) and (*2*) no. 2 ($\lambda = 791 \pm 8$ nm). The solid curves are fittings by the expression $\sigma(0) - \sigma(t) = A\ln(1 + t/\tau)$ for $\tau = 116$ s (*1*) and 41 s (*2*). The inset shows the dependence $\tau(T)$ for $\lambda \geq 1120$ nm.

QD layers is 18 nm) (this process cannot be detected on the time scale of our measurements). For holes, the δ-layers serve as a peculiar water divide. The holes will slide down from the external (with respect to the QD layer) part of a sample to the substrate and to the surface, and from the internal part of a sample—to the QD layer, where they recombine with electrons. Therefore, in the QD layer the negative charge is accumulated, while in the substrate and near the sample surface, the charge increases compared to the equilibrium situation. This will occur until the charge redistribution "straightens" the band diagram (Fig. 11b), after which photogenerated electrons and holes will cease to separate.

When filter no. 1 ($\lambda \geq 1120$ nm, $\varepsilon_\gamma < 1.1$ eV) is used, the photons can excite electron neither from the valence band nor from the hole levels in QD (the luminescence bands of our samples located in the region between 1.35 and 1.47 eV correspond to the transitions between the neighboring electronic and hole levels). The concentration increases due to excitation of electrons from the level of Cr in the substrate. In this case, of course, the reverse process will take place—the neutralization of charged donors by photoexcited electrons. The bands are inclined near the substrate–buffer layer interface (Fig. 11a), and the excited electrons begin move away from the substrate and can manage to slide down to the QD layer. This process of saturation of the QD layer by electrons occurs substantially slower than upon illumination by light with the energy exceeding the energy gap because in this case the electron concentration in the QD layer increases only due to a narrow layer adjacent to the substrate–GaAs buffer layer interface. In addition, the light transmitted by filter no. 1 (in contrast to filter no. 2) cannot neutralize shallow acceptors, which are always present in GaAs. When the QD layer is saturated by electrons, the band diagram becomes "straight" (as upon illumination through filter no. 2), but because of the presence of charged acceptors, the bands become completely straight at a smaller change in the electron concentration in the QD layer.

The described difference in the saturation times and the final concentration was observed in experiments. Upon illumination through filter no. 2, the conductivity saturated within several seconds, whereas upon illumination through filter no. 1 at the same intensity, the conductivity saturated within approximately half an hour.

The additional concentration of charge carriers required for straightening the bands in the GaAs buffer layer between the substrate and the QD layer is approximately equal to

$$\Delta n_s = \frac{\varepsilon_0 \varepsilon}{ed} \Delta V,$$

where $\Delta V \approx 0.6$ V is the potential corresponding to the energy gap between the chromium level in the substrate [29] and the Fermi energy in the QD layer (Fig. 11b); $d = 0.45$ μm is the buffer layer thickness in the structures studied; and $\varepsilon = 13.2$ is the dielectric constant of GaAs at 4.2 K. The value of Δn_s obtained in this way is 10^{11} cm^{-2}. Approximately the same number of electrons enter the QD layer from the free surface of the sample upon its illumination because in the equilibrium situation the Fermi level is located approximately by $\Delta V = 0.7$ V below the bottom of the conductivity band [30]. Therefore, a total increase in the electron concentration in the QD layer due to the band straightening can be estimated as 2×10^{11} cm^{-2}, which corresponds to the increase in the electron concentrations observed in samples after their illumination through filter no. 1, which were measured based on the Shubnikov–de Haas effect (see table). A small increase in the electron concentration in QD after illumination by light transmitted by filter no. 2 (compared to illumination through filter no. 1) is most likely explained by the neutralization of charged acceptors, which are present in small amounts in *i*-GaAs.

According to the above discussion, in our case the effect of persistent conductivity is related to the spatial separation of charges, the electrons entering QD layers after illumination of the both types. The spatial separation of charge carriers is also confirmed by the logarithmic dependence of the photoconductivity relaxation on time. Figure 12 presents the experimental data and their fit using the expression [31]

$$\sigma(0) - \sigma(t) = A\ln\left(1 + \frac{t}{\tau}\right), \qquad (2)$$

which is valid for the initial time interval. The neutralization of charged acceptors upon illumination through filter no. 2 results in a faster relaxation of the positive persistent photoconductivity than upon illumination

through filter no. 1 because of the recombination of electrons from the QD layer with neighboring acceptors. The parameter τ (see inset in Fig. 12) decreases with increasing temperature. This decrease is probably explained by the fact that thermally activated electrons more rapidly pass over the potential barrier in the GaAs buffer layer between the QD layer and the substrate and between QD and the sample surface.

4. CONCLUSION

We have studied the conductivity along QD layers in the InAs/GaAs structures. Depending on the concentration of current carriers, QD systems demonstrated both metallic properties (the Shubnikov–de Haas effect) and strong localization of current carriers. The passage from the metallic to hopping conductivity is caused by the increase in the effect of inhomogeneities of the potential relief of QD layers with decreasing electron concentration. The temperature dependence of the resistance of samples of the *p*-type with a low concentration of carriers changes at low temperatures from the dependence obeying the Mott law for the variable range hopping conductivity to the Shklovskii–Efros law for the hopping conductivity in the presence of the Coulomb gap in the density of states. Illumination of the structures results in an increase in the concentration of carriers in QD layers and enhances their mobility. The positive persistent conductivity observed in all the samples is caused by the spatial separation of photogenerated charges.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, projects nos. 00-02-17493 and 01-02-16441.

REFERENCES

1. N. N. Ledentsov, P. D. Wang, C. M. Sotomayor Torres, *et al.*, Phys. Rev. B **50**, 12171 (1994).
2. M. Grundman, O. Stier, and D. Bimberg, Phys. Rev. B **52**, 11969 (1995).
3. D. Bimberg, M. Grundman, and N. N. Ledentsov, *QD Heterostructures* (Wiley, West Sussex, 1998).
4. A. D. Yoffe, Adv. Phys. **50**, 1 (2001).
5. P. W. Fry, I. E. Itskevich, D. J. Mowbray, *et al.*, Phys. Rev. Lett. **84**, 733 (2000).
6. G. H. Kim, J. T. Nicholls, S. I. Khondaker, *et al.*, Phys. Rev. B **61**, 10910 (2000).
7. A. Hartmann, Y. Ducommun, E. Kapon, *et al.*, Phys. Rev. Lett. **84**, 5648 (2000).
8. K. Leosson, J. R. Jensen, and W. Langbein, Phys. Status Solidi B **221**, 49 (2000).
9. P. B. Joyce, T. J. Krzyzewski, G. R. Bell, *et al.*, Phys. Rev. B **58**, R15981 (1998).
10. S. Taddei, M. Colocci, and G. Salvati, Phys. Rev. B **62**, 10220 (2000).
11. O. Stier, M. Grundman, and D. Bimberg, Phys. Rev. B **59**, 5688 (1999).
12. H. L. Wang, F. H. Yang, S. L. Feng, *et al.*, Phys. Rev. B **61**, 5530 (2000).
13. L. Jacak, Eur. J. Phys. **21**, 487 (2000).
14. B. N. Zvonkov, E. R. Lin'kova, I. G. Malkina, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **63**, 418 (1996) [JETP Lett. **63**, 439 (1996)].
15. M. Kitamura, M. Nishioka, J. Oshinowo, and Y. Arakawa, Appl. Phys. Lett. **66**, 3663 (1995).
16. E. Ribeiro, E. Muller, T. Heinzel, *et al.*, Phys. Rev. B **58**, 1506 (1998).
17. A. J. Shields, M. P. O'Sullivan, I. Farrer, *et al.*, Appl. Phys. Lett. **74**, 735 (1999).
18. J. Y. Marzin, J. M. Gerard, A. Izrael, *et al.*, Phys. Rev. Lett. **73**, 716 (1994).
19. N. N. Ledentsov, V. A. Shchukin, M. Grundman, *et al.*, Phys. Rev. B **54**, 8743 (1996).
20. V. A. Kul'bachinskii, V. G. Kytin, R. A. Lunin, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **33**, 316 (1999) [Semiconductors **33**, 318 (1999)].
21. A. D. Visser, V. I. Kadushkin, V. A. Kul'bachinskii, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **59**, 339 (1994) [JETP Lett. **59**, 363 (1994)].
22. V. A. Kul'bachinskii, V. G. Kytin, V. I. Kadushkin, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **37**, 2693 (1995) [Phys. Solid State **37**, 1481 (1995)].
23. B. I. Shklovskii and A. L. Efros, *Electronic Properties of Doped Semiconductors* (Nauka, Moscow, 1979; Springer-Verlag, New York, 1984).
24. S. I. Khondaker, I. S. Shlimak, J. T. Nicholls, *et al.*, Phys. Rev. B **59**, 4580 (1999).
25. Yu. M. Gal'perin, E. P. German, and V. G. Karpov, Zh. Éksp. Teor. Fiz. **99**, 343 (1991) [Sov. Phys. JETP **72**, 193 (1991)].
26. S. Kivelson, D.-H. Lee, and S.-C. Zhang, Phys. Rev. B **46**, 2223 (1992).
27. C. E. Johnson and H. W. Jiang, Phys. Rev. B **48**, 2823 (1993).
28. *Delta-Doping of Semiconductors*, Ed. by E. F. Shubert (Cambridge Univ. Press, Cambridge, 1996).
29. B. Clerjaud, A. M. Hennel, and G. Martinez, Solid State Commun. **33**, 983 (1980).
30. V. V. Valyaev, V. L. Gurtovoi, D. Yu. Ivanov, *et al.*, Zh. Éksp. Teor. Fiz. **113**, 693 (1998) [JETP **86**, 383 (1998)].
31. H. J. Queisser and D. E. Theodorou, Phys. Rev. B **33**, 4027 (1986).

Translated by M. Sapozhnikov

Journal of Experimental and Theoretical Physics, Vol. 93, No. 4, 2001, pp. 824–832.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 945–953.

SOLIDS
Electronic Properties

Partial Moments of Electric Field Strength in the Problem of Conduction of Binary Composites

B. Ya. Balagurov

Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, ul. Kosygina 4, Moscow, 117997 Russia
e-mail: balagurov@deom.chph.ras.ru

Received March 6, 2001

Abstract—Some general properties of partial (calculated over the volume of individual components) moments of the electric field strength are considered in the standard linear problem of electrical conductivity of two-component media. The description of the critical behavior of moments of an arbitrary order in the vicinity of the metal–insulator phase transition point given here is free of any model assumptions. The corresponding critical indices are introduced and the relations between them following from the similarity hypothesis are established. The reciprocity relations obtained for 2D systems and connecting partial moments of the same order reduce by half the number of new independent critical indices. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In the theory of transport phenomena in two-component media, partial (averaged over the volume of individual components) characteristics of fields such as various moments of the electric field strength play an important role (see, for example, [1–4]). For instance, the effective electrical conductivity of such a medium can be expressed in terms of the first- as well as second-order moments. Structural field and current fluctuations [1, 2] and also the Joule heat liberated in each component can be expressed through second-order partial moments. The knowledge of second-order moments also make it possible to determine the derivatives of the effective conductivity σ_e with respect to its arguments, i.e., the conductivities σ_i of the components [2]. The corresponding relations make it possible to study comprehensively the critical behavior of effective conductivity (see, for example, [5]). It should be noted that the derivatives of σ_e also appear in the expressions for low- frequency permittivity [6], magnetoresistance in a weak magnetic field [2], and also (for a certain relation between parameters) in the expression for thermo-emf [7].

An analysis of nonlinear phenomena in inhomogeneous media necessitated the study of higher-order moments. For example, in order to calculate the first nonlinear correction to effective conductivity, the knowledge of fourth- order partial moments of the electric field strength is essential [3, 4]. Moments of sixth, eighth, etc. order appear in the next approximations in nonlinearity (see Appendix A). Fourth-order moments also appear in the problem of low-frequency noise spectrum in an inhomogeneous medium [3, 4]. It should be noted that some other quantities might be required in certain problems. For example, quadratic averages of the longitudinal and transverse components of the electric field strength appear in the problem of galvanomagnetic properties of two-component media in a weak magnetic field [2]. Finally, partial moments of odd orders are also of interest.

Thus, various effective characteristics of two-component media can be expressed in terms of partial moments of electric field strength of various orders. It should be noted that these moments, which are defined in the linear problem of electrical conduction, are functions of two arguments: concentration p and the conductivity ratio $h = \sigma_2/\sigma_1$ of the components. Consequently, the effective characteristics of the medium, which are initially functions of many parameters, can be reduced to the level of two-parametric functions. At the same time, the critical behavior of two-parametric functions in the vicinity of the metal–insulator phase-transition point can be described on the basis of the standard similarity hypothesis (cf. a similar procedure in the case of electrical conductivity [8] and some other quantities [2, 5]). This circumstance, in turn, makes it possible to describe consistently the critical behavior of multiparametric effective characteristics of the medium in the spirit of the conventional similarity hypothesis.

In the present work, we consider some general properties of partial moments of the electric field strength determined in the linear problem on the conduction of binary composites. The main attention is paid to the analysis of moments in the vicinity of the metal–insulator phase transition point. The critical behavior of the moments of an arbitrary order are described for two-component media; the corresponding indices are introduced and the relation following from the similarity hypothesis and connecting these indices are derived. It is shown that each partial moment of an order higher than two is characterized by a new critical index (as compared to the conduction indices). The reciprocity

relations found for two-dimensional systems make it possible to relate the partial moments of the same order, which are calculated for the first and second components. As a consequence, the number of new indices in the 2D case is half as large as in the 3D case.

The algorithm of consistent computation of the effective nonlinear conductivity for a nonhomogeneous medium is given in Appendix A. In Appendix B, a concise derivation of the formula for the noise spectrum (low-frequency current fluctuations) in a binary composite, generalizing to a certain extent the conventional derivation [3, 4], is presented.

2. EFFECTIVE CHARACTERISTICS OF A MEDIUM

The problem of the conductivity of a nonhomogeneous isotropic medium is formulated in the following standard way. The system of equations for a direct current can be written in the form

$$\mathrm{curl}\mathbf{E} = 0, \quad \mathrm{div}\mathbf{j} = 0, \tag{1}$$

where $\mathbf{E}$ is the electric field strength and $\mathbf{j}$ is the current density. In the approximation linear in the electric field, the quantities $\mathbf{j}$ and $\mathbf{E}$ are connected through Ohm's law:

$$\mathbf{j} = \sigma(\mathbf{r})\mathbf{E}, \tag{2}$$

where $\sigma(\mathbf{r})$ is the coordinate-dependent local conductivity of the medium. The system of equations (1) and (2) is solved under the condition that a uniform field $\langle\mathbf{E}\rangle$ exists in the medium, where $\langle\ldots\rangle$ denotes averaging over the volume V of the sample. The effective conductivity σ_e of the medium is defined as the proportionality factor in the expression connecting the average current density $\langle\mathbf{j}\rangle$ and the quantity $\langle\mathbf{E}\rangle$:

$$\langle\mathbf{j}\rangle = \sigma_e\langle\mathbf{E}\rangle \tag{3}$$

for $V \longrightarrow \infty$. Thus, conductivity σ_e is the linear response of the medium to the applied external "perturbation," viz., the uniform field $\langle\mathbf{E}\rangle$.

The effective conductivity σ_e can be expressed in terms of quadratic parameters of the field. By virtue of Eqs. (1), the following well-known identity holds (see, for example, [1, 2]):

$$\langle\mathbf{jE}\rangle = \langle\mathbf{j}\rangle\langle\mathbf{E}\rangle, \tag{4}$$

which is valid for an arbitrary (including nonlinear) dependence of $\mathbf{j}$ on $\mathbf{E}$. In the linear case, Eq. (4) combined with (2) and (3) gives

$$\sigma_e = \langle\sigma\mathbf{e}^2\rangle, \tag{5}$$

where

$$\mathbf{e}(\mathbf{r}) = \frac{\mathbf{E}(\mathbf{r})}{|\langle\mathbf{E}\rangle|} \tag{6}$$

is the dimensionless electric field strength in the medium.

For a two-component medium, the quantity $\sigma(\mathbf{r})$ assumes the constant values σ_1 and σ_2 for the first and second component, respectively. In this case, Eq. (5) leads to

$$\sigma_e = \sigma_1\psi_1^{(2)} + \sigma_2\psi_2^{(2)}. \tag{7}$$

Here

$$\psi_i^{(2)} = \langle\mathbf{e}^2\rangle^{(i)} \quad (i = 1, 2) \tag{8}$$

and

$$\langle(\ldots)\rangle^{(i)} = \frac{1}{V}\int_{V_i}(\ldots)d\mathbf{r}, \tag{9}$$

where integration is carried out over the volume V_i of the ith component. Quadratic quantities $\psi_i^{(2)}$ will be referred to as second-order partial moments of the electric field strength. It should be emphasized that, like conductivity σ_e, functions $\psi_i^{(2)}$ are effective characteristics of the medium (self-averaging for $V \longrightarrow \infty$) and are independent of the magnitude and direction (isotropic case) of the applied field $\langle\mathbf{E}\rangle$.

For the two-component systems under investigation, it is convenient to introduce the dimensionless effective conductivity f in accordance with the relations

$$\sigma_e = \sigma_e(p; \sigma_1, \sigma_2) \equiv \sigma_1 f(p, h), \quad h = \frac{\sigma_2}{\sigma_1}, \tag{10}$$

where p is the concentration (a fraction of the occupied volume) of the first component. In this case, relation (7) leads to

$$f = \psi_1^{(2)} + h\psi_2^{(2)}. \tag{11}$$

It should be noted that the quantities $\psi_i^{(2)}$ are determined only by the properties of the medium and are functions of the same arguments as function f:

$$\psi_i^{(2)} = \psi_i^{(2)}(p, h).$$

The role of functions $\psi_i^{(2)}$ is not limited to their use in relations (7) and (11). For example, the quadratic structural fluctuations of fields and currents (see, for example, [1, 2])

$$\Delta_E^2 = \langle(\mathbf{E} - \langle\mathbf{E}\rangle)^2\rangle(\langle\mathbf{E}\rangle)^{-2},$$

$$\Delta_j^2 = \langle(\mathbf{j} - \langle\mathbf{j}\rangle)^2\rangle(\langle\mathbf{j}\rangle)^{-2}$$

can also be expressed in terms of the quantities $\psi_i^{(2)}$:

$$\Delta_E^{(2)} = \sum_i \psi_i^{(2)} - 1,$$
$$\Delta_j^{(2)} = \frac{1}{\sigma_e^2}\sum_i \sigma_i^2 \psi_i^{(2)} - 1. \quad (12)$$

The fraction of the Joule heat Q_i liberated per unit time in the volume of the ith component can also be expressed in terms of $\psi_i^{(2)}$:

$$Q_i = \langle \mathbf{j} \cdot \mathbf{E} \rangle^{(i)} = \sigma_i \psi_i^{(2)} (\langle \mathbf{E} \rangle)^2. \quad (13)$$

Using identities of type (4), we can relate functions $\psi_i^{(2)}$ with the derivatives of σ_e with respect to σ_i:

$$\psi_i^{(2)} = \langle \mathbf{e}^2 \rangle^{(i)} = \frac{\partial \sigma_e}{\partial \sigma_i}. \quad (14)$$

Substituting relations (10) into (14), we obtain [2]

$$\psi_1^{(2)} = f - h f', \quad \psi_2^{(2)} = f', \quad f \equiv \frac{\partial f(p, h)}{\partial h}. \quad (15)$$

These relations make it possible to express the quantities $\Delta_E^{(2)}$, $\Delta_j^{(2)}$, and Q_i in terms of function f and its derivative f'. On the other hand, the application of relations (15) allows us to find the derivative f' in a numerical experiment without a cumbersome numerical differentiation (see [5]). It should be noted that the knowledge of the quantity $\psi_2^{(2)} = f'$ in the entire range of its argument is essential for determining the low-frequency permittivity [6], thermo-emf [7], and magnetoresistance [2]. Besides, when the quantities f and f' are determined together, it is possible to carry out a more detailed analysis of the critical behavior of conductivity (see [5]).

It was proved in [3, 4] that the calculation of high-order moments of the electric field strength (e.g., fourth-order moments) are required for studying nonlinear properties of nonhomogeneous media. For a weakly nonlinear isotropic medium, instead of relation (2) we have

$$\mathbf{j} = \{\sigma(\mathbf{r}) + \chi^{(3)}(\mathbf{r})\mathbf{E}^2 + \chi^{(5)}(\mathbf{r})\mathbf{E}^4 + \ldots\}\mathbf{E}. \quad (16)$$

The average values $\langle \mathbf{j} \rangle$ and $\langle \mathbf{E} \rangle$ are connected through a similar relation:

$$\langle \mathbf{j} \rangle = \{\sigma_e + \chi_e^{(3)}(\langle \mathbf{E} \rangle)^2 + \chi_e^{(5)}(\langle \mathbf{E} \rangle)^4 + \ldots\}\langle \mathbf{E} \rangle, \quad (17)$$

where $\chi_e^{(3)}$, $\chi_e^{(5)}$, … are the effective nonlinearity coefficients. According to [3, 4] (see also Appendix A), the quantity $\chi_e^{(3)}$ can be expressed in terms of the electric field strength in the linear problem:

$$\chi_e^{(3)} = \langle \chi^{(3)} \mathbf{e}^4 \rangle, \quad (18)$$

where $\mathbf{e}(\mathbf{r})$ is the same as in relation (6). For a two-component medium, it follows from (18) that

$$\chi_e^{(3)} = \chi_1^{(3)} \psi_1^{(4)} + \chi_2^{(3)} \psi_2^{(4)}, \quad (19)$$

where $\chi_i^{(3)}$ is the value of the nonlinearity coefficient $\chi^{(3)}(\mathbf{r})$ for the ith component and

$$\psi_i^{(4)} = \langle \mathbf{e}^4 \rangle^{(i)} \quad (20)$$

is the fourth-order partial moment. Obviously, the value of $\psi_i^{(4)}$ is determined only by the properties of the medium and is a function of the same arguments as f and $\psi_i^{(2)}$:

$$\psi_i^{(4)} = \psi_i^{(4)}(p, h).$$

Fourth-order moments also appear in the problem of low-frequency noise spectrum in nonhomogeneous samples. According to [3, 4] (see also Appendix B), the noise spectrum $K(\omega)$ can be expressed in terms of the electric field strength $\mathbf{e}(\mathbf{r})$ in the linear problem:

$$K(\omega) = \frac{1}{V}\langle \lambda_\omega \mathbf{e}^4 \rangle \sigma_e^{-2}. \quad (21)$$

Here, $K(\omega)$ and $\lambda_\omega(\mathbf{r})$ are the Fourier transforms of the functions $K(\tau)$ and $\lambda(\tau, \mathbf{r})$ defined in Appendix B. For a two-component medium, relation (21) leads to

$$K(\omega) = \frac{1}{V}[\lambda_1(\omega)\psi_1^{(4)} + \lambda_2(\omega)\psi_2^{(4)}]\sigma_e^{-2}, \quad (22)$$

where $\lambda_i(\omega)$ is the value of the quantity $\lambda_\omega(\mathbf{r})$ for the ith component and $\psi_i^{(4)}$ is the same as in relation (20).

It follows from Eqs. (19) and (22) that for studying the behavior of the quantities $\chi_e^{(3)}$ and $K(\omega)$ in the vicinity of the metal–insulator phase-transition point, it is sufficient to analyze the critical behavior of the partial moments $\psi_i^{(4)}$. In this case, the ratios of parameters $\chi_2^{(3)}/\chi_1^{(3)}$ and $\lambda_2(\omega)/\lambda_1(\omega)$ are generally arbitrary so that any of the two terms in expressions (19) and (22) may dominate.

Thus, in order to determine various physical effective parameters of binary composites, it is sufficient to calculate the partial moments

$$\psi_i^{(2n)} = \langle \mathbf{e}^{2n} \rangle^{(i)} \quad (n = 1, 2, \ldots) \quad (23)$$

with $\mathbf{e}(\mathbf{r})$ from (6). This necessitates an analysis of the quantity $\psi_i^{(2n)} = \psi_i^{(2n)}(p, h)$ in the entire range of arguments p and h, including the neighborhood of the metal–insulator phase-transition point.

It would be also interesting to analyze odd-order partial moments $\psi_i^{(2n+1)}(p, h)$ defined by the relation

$$\langle \mathbf{e}^{2n+1} \rangle^{(i)} = \psi_i^{(2n+1)} \langle \mathbf{e} \rangle. \quad (24)$$

Here, we take into account the fact that the vector quantity $\langle e^{2n}\mathbf{e}\rangle^{(i)}$ for an isotropic medium can be directed only along the unit vector $\langle \mathbf{e} \rangle = \langle \mathbf{E} \rangle / |\langle \mathbf{E} \rangle|$. It follows from Eq. (24) that

$$\psi_i^{(2n+1)} = \langle e^{2n} e_\parallel \rangle^{(i)}, \quad (25)$$

where $e_\parallel(\mathbf{r})$ is the component of $\mathbf{e}(\mathbf{r})$ parallel to $\langle \mathbf{E} \rangle$. It should be noted that, in accordance with Eq. (3), the effective conductivity can be expressed in terms of the first-order partial moments $\psi_i^{(1)}$:

$$\sigma_e = \sigma_1 \psi_1^{(1)} + \sigma_2 \psi_2^{(1)}. \quad (26)$$

The computation of odd higher-order moments is required, for example, for analyzing structural fluctuations of the electric field strength of the form

$$\Delta_E^{(2n)} = \langle (\mathbf{e} - \langle \mathbf{e} \rangle)^{2n} \rangle \quad (27)$$

with $n \geq 2$.

3. CRITICAL BEHAVIOR OF FUNCTIONS $\psi_i^{(2n)}(p, h)$

The critical behavior of the effective conductivity of a randomly nonhomogeneous two-component medium in the vicinity of the metal–insulator phase-transition point can be described using the similarity hypothesis [8]. According to [8] (see also, for example, [2]), function f in the critical region $h \ll 1$, $|\tau| \ll 1$, where $\tau = (p - p_c)/p_c$, p_c being the critical concentration) varies as follows:

$$f = \tau^t \left\{ A_0 + A_1 \frac{h}{\tau^{t/s}} + \ldots \right\}, \quad (28)$$

for $\tau > 0$, $\Delta_0 \ll \tau \ll 1$

$$f = h^s \left\{ a_0 + a_1 \frac{\tau}{h^{s/t}} + \ldots \right\}, \quad (29)$$

for $|\tau| \ll \Delta_0$, and

$$f = \frac{h}{(-\tau)^q} \left\{ B_1 + B_2 \frac{h}{(-\tau)^{t/s}} + \ldots \right\}, \quad (30)$$

for $\tau < 0$, $\Delta_0 \ll |\tau| \ll 1$. Here

$$\Delta_0 = h^{s/t} \quad (31)$$

is the size of the "smearing" region [8]. The critical indices t, q, and s are connected through the following relation:

$$q = \frac{t}{s} - t. \quad (32)$$

Numerical experiments on disordered lattices (see, for example, [3, 5, 9, 10]) in the 3D case give

$$t \approx 2, \quad q \approx 0.8, \quad s \approx 0.7. \quad (33)$$

It is natural to expect that the basic concepts of the similarity hypothesis, including expressions (28)–(33), are also applicable to finely dispersed binary composites.

The critical behavior of functions $\psi_i^{(2)}$ (see [5]) can be revealed using relations (15) after the substitution of expressions (28)–(30) into them. All the critical indices of quantities $\psi_i^{(2)}$ in this case can be expressed in terms of the conductivity indices t, q, and s (see (42)). At the same time, no relations of type (15) which would connect $\psi_i^{(2n)}$ with the dimensionless effective conductivity f are known for functions (23) with $n \geq 2$. For this reason, the behavior of the quantities $\psi_i^{(2n)}(p, h)$ with $n \geq 2$ in the vicinity of the metal–insulator phase-transition point should be determined proceeding from their general properties. Such a procedure was carried out, for instance, in [2, 5] for the two-parametric functions emerging in the problem of galvanomagnetic properties of binary systems in a weak magnetic field.

Confining the analysis to the principal terms in the corresponding expansions and omitting factors of the order of unity, in analogy with [2, 5], we obtain the following expressions for function $\psi_i^{(2n)}(p, h)$ in the critical region:

$$\psi_1^{(2n)} \sim \tau^{t_{2n}}, \quad \psi_2^{(2n)} \sim \tau^{-\mu_{2n}}, \quad (34)$$

for $\tau > 0$, $\Delta_0 \ll \tau \ll 1$,

$$\psi_1^{(2n)} \sim h^{s_{2n}}, \quad \psi_2^{(2n)} \sim h^{-\lambda_{2n}}, \quad (35)$$

for $|\tau| \ll \Delta_0$, and

$$\psi_1^{(2n)} \sim \frac{h^{2n}}{(-\tau)^{q_{2n}}}, \quad \psi_2^{(2n)} \sim \frac{1}{(-\tau)^{\mu_{2n}}} \quad (36)$$

for $\tau < 0$, $\Delta_0 \ll |\tau| \ll 1$. Here

$$q_{2n} = 2n\frac{t}{s} - t_{2n}, \quad s_{2n} = \frac{s}{t} t_{2n}, \quad \lambda_{2n} = \frac{s}{t} \mu_{2n}. \quad (37)$$

By virtue of relations (37), only two of the five new critical indices (for a fixed n) are independent (for these indices, we can choose, for example, t_{2n} and μ_{2n}).

Expressions (34)–(37) were obtained on the basis of the following considerations. In the concentration range $\tau > 0$, $\Delta_0 \ll \tau \ll 1$, the electric field is "expelled" from the high-conductivity (first) component as we

approach the transition point (cf. similar considerations in [2]). Consequently, $\psi_1^{(2n)} \longrightarrow 0$ for $\tau \longrightarrow 0$. In accordance with the similarity hypothesis, we assume that this decrease follows a power law. At the same time, the quantity $\psi_2^{(2n)}$ in the given case increases since the number of narrow "bridges" (junctions) formed by the high-conductivity component increases. In the vicinity of these junctions, the electric field strength in the low-conductivity (second) component increases sharply, which is reflected in formula (34) for $\psi_2^{(2n)}$.

For $p < p_c$ and $\sigma_2 = 0$ (i.e., for $\tau < 0$ and $h = 0$), the electric field strength in the inclusions of the first component forming finite clusters is equal to zero. For small values of $h \neq 0$, the field strength in the high-conductivity component differs from zero, but is also small and is linear in h outside the smearing region (cf. [2]). Consequently, function $\psi_1^{(2n)}$ in relations (36) is proportional to h^{2n}. On the other hand, for a fixed h, the value of $\psi_1^{(2n)}$ must increase for $p \longrightarrow p_c$, which is taken into account in formula (36). For $p < p_c$ and $p \longrightarrow p_c$, function $\psi_2^{(2n)}$ also increases since the number of "hot" points near which the electric field strength increases sharply becomes larger in this case. Finally, formulas (35) give the values of $\psi_1^{(2n)}$ and $\psi_2^{(2n)}$ in the smearing region, in which finite (nonzero) conductivities of both components must be taken into account.

The smearing region $\Delta_1^{(2n)}$ for function $\psi_1^{(2n)}$ is determined in the conventional way by joining expressions (34) and (35):

$$\Delta_1^{(2n)} = h^{s_{2n}/t_{2n}}.$$

Equating now expressions (35) and (36) for $\psi_1^{(2n)}$ for $\tau = -\Delta_1^{(2n)}$, we arrive at the first relation for the critical indices t_{2n}, s_{2n}, and q_{2n}. Further, in accordance with the similarity hypothesis, all the critical indices in the problem of electrical conductivity must be characterized by the universal scale so that $\Delta_1^{(2n)} \sim \Delta_0$, where Δ_0 is defined in (31). This leads to the second relation from (37), which was taken into account while writing the first relation.

The quantity $\psi_2^{(2n)}$ is considered similarly. In this case, the equality of critical indices for $\tau > 0$ and $\tau < 0$ follows from the "joining" of relations (34) and (36) with (35), which is symmetric relative to $\tau = 0$. Accordingly, we derive the relation between indices λ_{2n} and μ_{2n} (third relation in (37)) by equating (by the order of magnitude) the size $\Delta_2^{(2n)}$ of the smearing region for function $\psi_2^{(2n)}$ and the quantity Δ_0 from relation (31).

It should be noted that

$$\begin{aligned}&\left\langle \left(\mathbf{e}^{2n} - \left\langle \mathbf{e}^{2n}\right\rangle^{(i)}\right)^2\right\rangle^{(i)}\\ &= \psi_i^{(4n)} - (2 - p_i)\left(\psi_i^{(2n)}\right)^2 \geq 0,\end{aligned} \tag{38}$$

where p_i is the concentration of the ith component ($p_1 = p$, $p_2 = 1 - p$). Substituting the expressions for $\psi_i^{(4n)}$ and $\psi_i^{(2n)}$ from formulas (34)–(36) into relation (38), we obtain a number of inequalities for critical indices, from which the following two inequalities are independent:

$$t_{4n} \leq 2t_{2n}, \quad \mu_{4n} \geq 2\mu_{2n}. \tag{39}$$

The remaining inequalities can be reduced to these two inequalities when relations (37) are taken into consideration.

For a randomly inhomogeneous medium, the simultaneous substitutions $\sigma_1 \rightleftharpoons \sigma_2$ and $p \longrightarrow 1 - p$ do not change the properties of the medium as a single entity (see [1]) so that $\sigma_e(p; \sigma_1, \sigma_2) = \sigma_e(1 - p; \sigma_2, \sigma_1)$, whence

$$f(p, h) = hf\left(1 - p, \frac{1}{h}\right). \tag{40}$$

Accordingly, for partial moments $\psi_i^{(2n)}$, we have

$$\begin{aligned}\psi_1^{(2n)}\left(1 - p, \frac{1}{h}\right) &= \psi_2^{(2n)}(p, h),\\ \psi_2^{(2n)}\left(1 - p, \frac{1}{h}\right) &= \psi_1^{(2n)}(p, h).\end{aligned} \tag{41}$$

These equalities are quite obvious since the first and second components change places under the substitutions $\sigma_1 \rightleftharpoons \sigma_2$ and $p \longrightarrow 1 - p$. Relations (41) make it possible to determine the quantities $\psi_i^{(2n)}(p, h)$ for $h > 1$ if they are known for $h < 1$ in the entire concentration range.

It was noted above that no new independent indices appear for the second-order moments $\psi_i^{(2)}$. Indeed, substituting expressions (28)–(30) into relation (15) and comparing the result with (34)–(36) (for $n = 1$), we obtain

$$\begin{aligned}t_2 = t, \quad s_2 = s, \quad q_2 = q + \frac{t}{s};\\ \mu_2 = q, \quad \lambda_2 = 1 - s.\end{aligned} \tag{42}$$

It can be easily verified that indices (42) satisfy relations (37) for $n = 1$. The quantity λ_2 is positive since $s < 1$, which is in turn a consequence of the positiveness of function $\psi_1^{(2)}$.

In the case of fourth-order moments, critical indices are usually introduced for functions $\chi_e^{(3)}/\sigma_e^2$ (see [3, 4]).

In our notation, the following relations correspond to the relevant relations from [3, 4]:

$$\frac{\psi_1^{(4)}}{f^2} \sim \frac{1}{\tau^k}, \tag{43}$$

for $\tau > 0$, $\Delta_0 \ll \tau \ll 1$ and

$$\frac{h^2\psi_2^{(4)}}{f^2} \sim \frac{1}{(-\tau)^{k'}} \tag{44}$$

for $\tau < 0$, $\Delta_0 \ll |\tau| \ll 1$. The substitution of expressions from (28), (34) and (30), (36) into relations (43) and (44) gives

$$k = 2t - t_4, \quad k' = \mu_4 - 2q. \tag{45}$$

Inequalities (39) combined with definitions (45) and relations (42) give [3]

$$k \geq 0, \quad k' \geq 0. \tag{46}$$

If, following [10], we assume for estimates that $t = 2$ and $q = 0.75$ (in this case, it follows from relation (32) that $s \approx 0.73$), $k = 1.6$ and $k' = 0.7$ we obtain from relations (45) and (37) the following values of the remaining indices:

$$t_4 \approx 2.4, \quad \mu_4 \approx 2.2, \quad q_4 \approx 8.6,$$

$$s_4 \approx 0.9, \quad \lambda_4 \approx 0.8.$$

It should be noted, however, that the calculation of indices using formulas (45) and (37) leads to noticeable errors (this refers especially to the value of q_4).

The critical behavior of odd momenta $\psi_i^{(2n+1)}(p, h)$ can be analyzed exactly in the same way as that of $\psi_i^{(2n)}(p, h)$ and is defined by formulas (34)–(37) for $n \geq 1$, in which $2n$ should be replaced by $2n + 1$. For $n = 0$, the quantities $\psi_i^{(1)} = \langle e_\parallel \rangle^{(i)}$ can be expressed in terms of dimensionless effective conductivity [2]:

$$\psi_1^{(1)} = \frac{f - h}{1 - h}, \quad \psi_2^{(1)} = \frac{1 - f}{1 - h}. \tag{47}$$

For $h \ll 1$, we have $\psi_1^{(1)} \approx f$ so that the critical behavior of $\psi_1^{(1)}$ is defined by formulas (28)–(32). In the case of randomly inhomogeneous media, relations of type (41) are valid for functions $\psi_i^{(2n+1)}$ after the substitution of $2n + 1$ for $2n$.

4. TWO-DIMENSIONAL CASE

The previous analysis was valid for 3D as well as 2D systems. In the 2D case, however, the so-called reciprocity relations can be established for quantities $\psi_i^{(2n)}$, which provide additional information on the properties of these functions. The subsequent analysis is valid only for 2D systems and has no three-dimensional analogue.

Following [1], we transform the relations between the field and the current to the following system of equations:

$$\begin{aligned} j_x &= \lambda E_y', \quad j_y = -\lambda E_x', \\ E_x &= \frac{1}{\lambda} j_y', \quad E_y = -\frac{1}{\lambda} j_x'. \end{aligned} \tag{48}$$

Under transformation (48) (with the parameter λ independent of coordinates), Eqs. (1) and (2) for direct current preserve their form, and the conductivity of system (48) can be written as

$$\sigma'(\mathbf{r}) = \frac{\lambda^2}{\sigma(\mathbf{r})}. \tag{49}$$

For a two-component system, it is convenient to put $\lambda = \sqrt{\sigma_1\sigma_2}$ so that the conductivity components change places under transformation (48):

$$\sigma_1' = \sigma_2, \quad \sigma_2' = \sigma_1, \quad h' = \sigma_2'/\sigma_1' = 1/h.$$

Calculating the mean values $\langle \mathbf{j}' \rangle$ and $\langle \mathbf{E}' \rangle$, we obtain the reciprocity relations for the effective conductivity:

$$\sigma_e(p; \sigma_1, \sigma_2)\sigma_e(p; \sigma_2, \sigma_1) = \sigma_1\sigma_2. \tag{50}$$

Substituting relations (10) into (50), we arrive at the reciprocity relation for the dimensionless effective conductivity:

$$f(p, h) f\left(p, \frac{1}{h}\right) = 1. \tag{51}$$

Further, from relations (48) we obtain

$$(\mathbf{E}')^2 = \frac{1}{\lambda^2}\mathbf{j}^2 = \left(\frac{\sigma(\mathbf{r})}{\lambda}\right)^2 \mathbf{E}^2, \tag{52}$$

$$(\langle \mathbf{E}' \rangle)^2 = \frac{1}{\lambda^2}(\langle \mathbf{j} \rangle)^2 = \left(\frac{\sigma_e}{\lambda}\right)^2 (\langle \mathbf{E} \rangle)^2, \tag{53}$$

so that

$$(\mathbf{e}')^2 = \left(\frac{\sigma(\mathbf{r})}{\sigma_e}\right)^2 \mathbf{e}^2. \tag{54}$$

Using expression (54), we obtain the following relations for the two-component system:

$$\psi_1^{(2n)}\left(p, \frac{1}{h}\right) = \frac{1}{[f(p, h)]^{2n}} \psi_1^{(2n)}(p, h), \tag{55}$$

$$\psi_2^{(2n)}\left(p, \frac{1}{h}\right) = \left[\frac{h}{f(p, h)}\right]^{2n} \psi_2^{(2n)}(p, h). \tag{56}$$

While deriving these formulas, we have assumed that $h' = 1/h$. Equalities (55) and (56) are the required reciprocity relations for functions $\psi_i^{(2n)}$.

Relations (50), (51), (55), and (56) are valid for isotropic two-dimensional two-component systems of an arbitrary structure (both periodic and disordered). Let us now consider a randomly inhomogeneous medium. Taking into account equalities (40) and (41), we can write relations (51), (55), and (56) for such a medium in the form

$$f(p, h)\, f(1 - p, h) = h, \tag{57}$$

$$\psi_1^{(2n)}(1 - p, h) = \left[\frac{h}{f(p, h)}\right]^{2n} \psi_2^{(2n)}(p, h), \tag{58}$$

$$\psi_2^{(2n)}(1 - p, h) = \frac{h}{[f(p, h)]^{2n}} \psi_1^{(2n)}(p, h). \tag{59}$$

Equalities (58) and (59) are transformed into each other after the substitution $p \longrightarrow 1 - p$.

For $p = p_c = 1/2$, relation (57) leads to the well-known result obtained by Dykhne [1]:

$$f\left(\frac{1}{2}, h\right) = \sqrt{h}. \tag{60}$$

In this case, we obtain the following relation from (58) (or (59)):

$$\psi_1^{(2n)}\left(\frac{1}{2}, h\right) = h^n \psi_2^{(2n)}\left(\frac{1}{2}, h\right), \tag{61}$$

which can also be written for $p = 1/2$ in the form

$$\langle (\mathbf{j} \cdot \mathbf{E})^n \rangle^{(1)} = \langle (\mathbf{j} \cdot \mathbf{E})^n \rangle^{(2)}. \tag{62}$$

This equality was derived earlier in [1].

In accordance with relation (60), index $s = 1/2$ for a 2D randomly inhomogeneous system, and $q = t$ in accordance with relation (32). Consequently, relations (37) in the 2D case assume the form

$$\begin{gathered} q_{2n} = 4nt - t_{2n}, \quad s_{2n} = \frac{1}{2t} t_{2n}, \\ \lambda_{2n} = \frac{1}{2t} \mu_{2n}. \end{gathered} \tag{63}$$

Substituting expressions (35) into relation (61), we obtain one more relation between critical indices:

$$s_{2n} = n - \lambda_{2n}, \tag{64}$$

which, together with relations (63), leads to the following relation between the two indices chosen in Section 3 as independent indices:

$$t_{2n} = 2nt - \mu_{2n}. \tag{65}$$

Relations (58) and (59) for $p \neq 1/2$ give no new relations between the indices. Thus, in the 2D case, only one new independent critical index (e.g., μ_{2n}) appears for quantities $\psi_1^{(2n)}$ and $\psi_2^{(2n)}$ for a fixed n.

For $n = 2$, relation (65) can also be written in the form $2t - t_4 = \mu_4 - 2t$. A comparison of this equality with (45) (for $q = t$) gives

$$k = k'. \tag{66}$$

If we assume that $t = q = 1.3$ [3, 11] and $k = k' = 1.2$ [10] for our estimates, we obtain the following values for the remaining critical indices in the 2D case:

$$t_4 \approx 1.4, \quad \mu_4 \approx 3.8, \quad q_4 = 9.0,$$

$$s_4 \approx 0.54, \quad \lambda_4 \approx 1.46.$$

Taking into account considerable errors introduced when the indices are calculated using formulas (63)–(66), we can assume that the value of $\lambda_4 \approx 1.46$ is in satisfactory agreement with the value of $\lambda_4 = 1.33 \pm 0.05$ obtained in [12].

It should be noted in conclusion that the isomorphism relations obtained in [13, 14] for a 2D two-component system in a transverse magnetic field h make it possible to relate the electric field strength for $\mathbf{H} \neq 0$ with the field strength in the same system for $\mathbf{H} = 0$. Thus, the partial moments for $\mathbf{H} \neq 0$ can be expressed in terms of the corresponding moments for $\mathbf{H} = 0$. Consequently, the results obtained in this section make it possible to study the properties of partial moments as functions of the magnetic field $\mathbf{H}$ in the vicinity of the metal–insulator phase-transition point. Such an analysis was carried out in [15] for functions $\psi_i^{(2)}(\mathbf{H})$.

APPENDIX A

We will seek the solution of Eqs. (1) for direct current in the case of a weakly nonlinear medium with Ohm's law in the form (16) using perturbation theory with the help of a power expansion in $\langle \mathbf{E} \rangle$:

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^{(1)}(\mathbf{r}) + \mathbf{E}^{(3)}(\mathbf{r}) + \mathbf{E}^{(5)}(\mathbf{r}) + \ldots, \tag{A.1}$$

$$\mathbf{j}(\mathbf{r}) = \mathbf{j}^{(1)}(\mathbf{r}) + \mathbf{j}^{(3)}(\mathbf{r}) + \mathbf{j}^{(5)}(\mathbf{r}) + \ldots. \tag{A.2}$$

Here, $\mathbf{E}^{(n)}(\mathbf{r})$ and $\mathbf{j}^{(n)}(\mathbf{r})$ are terms of the order of $|\langle \mathbf{E} \rangle|^n$. In accordance with Eqs. (16) and (A.1), (A.2), we can write the following expressions for $\mathbf{j}^{(n)}(\mathbf{r})$:

$$\begin{gathered} \mathbf{j}^{(1)}(\mathbf{r}) = \sigma(\mathbf{r})\mathbf{E}^{(1)}(\mathbf{r}), \\ \mathbf{j}^{(3)}(\mathbf{r}) = \sigma(\mathbf{r})\mathbf{E}^{(3)}(\mathbf{r}) + \chi^{(3)}(\mathbf{r})(\mathbf{E}^{(1)}(\mathbf{r}))^3, \\ \mathbf{j}^{(5)}(\mathbf{r}) = \sigma(\mathbf{r})\mathbf{E}^{(5)}(\mathbf{r}) + 3\chi^{(3)}(\mathbf{r})(\mathbf{E}^{(1)}(\mathbf{r}))^2 \mathbf{E}^{(3)}(\mathbf{r}) \\ + \chi^{(5)}(\mathbf{r})(\mathbf{E}^{(1)}(\mathbf{r}))^5, \ldots. \end{gathered} \tag{A.3}$$

The corresponding potentials defined in accordance with $\mathbf{E}^{(n)}(\mathbf{r}) = -\nabla\varphi^{(n)}(\mathbf{r})$ obey the following equations:

$$\begin{gathered} \nabla\{\sigma(\mathbf{r})\nabla\varphi^{(1)}(\mathbf{r})\} = 0, \\ \nabla\{\sigma(\mathbf{r})\nabla\varphi^{(3)}(\mathbf{r})\} = \nabla\{\chi^{(3)}(\mathbf{r})(\mathbf{E}^{(1)}(\mathbf{r}))^3\}, \ldots, \end{gathered} \tag{A.4}$$

whose solution is assumed to be known. We also assume that the problem can be solved for a given potential difference (i.e., for a preset $\langle \mathbf{E} \rangle$) so that

$$\langle \mathbf{E}^{(1)} \rangle = \langle \mathbf{E} \rangle, \quad \langle \mathbf{E}^{(n)} \rangle = 0, \quad n = 3, 5, \ldots. \tag{A.5}$$

In order to calculate the values of $\chi_e^{(3)}$, $\chi_e^{(5)}$, …, we will apply an approach similar to that used in the problems of thermo-emf [7] and magnetoresistance [2]. It will be shown below that this approach simplifies computations to a certain extent.

It should be noted that $\mathbf{E}^{(m)}(\mathbf{r})$ and $\mathbf{j}^{(n)}(\mathbf{r})$ satisfy equations of type (1):

$$\mathrm{curl}\mathbf{E}^{(m)} = 0, \quad \mathrm{div}\mathbf{j}^{(n)} = 0. \tag{A.6}$$

Consequently, these quantities satisfy relations similar to Eq. (4):

$$\begin{aligned} \langle \mathbf{j}^{(n)} \cdot \mathbf{E}^{(m)} \rangle &= \langle \mathbf{j}^{(n)} \rangle \langle \mathbf{E}^{(m)} \rangle; \\ n = 1, 3, 5, \ldots; \quad m &= 1, 3, 5, \ldots. \end{aligned} \tag{A.7}$$

For $n = 3$, $m = 1$ and $n = 1$, $m = 3$, relation (A.7) leads to

$$\langle \mathbf{j}^{(3)} \cdot \mathbf{E}^{(1)} \rangle = \langle \mathbf{j}^{(3)} \rangle \langle \mathbf{E}^{(1)} \rangle, \quad \langle \mathbf{j}^{(1)} \cdot \mathbf{E}^{(3)} \rangle = 0, \tag{A.8}$$

where it is assumed that $\langle \mathbf{E}^{(3)} \rangle = 0$. Taking into account the first of these equalities, the fact that $\langle \mathbf{j}^{(3)} \rangle = \chi_e^{(3)} (\langle \mathbf{E} \rangle)^3$, and the definition of the quantity $\mathbf{j}^{(3))}(\mathbf{r})$ from Eqs. (A.3), we obtain

$$\chi_e^{(3)} (\langle \mathbf{E} \rangle)^4 = \langle \sigma \mathbf{E}^{(1)} \cdot \mathbf{E}^{(3)} \rangle + \langle \chi^{(3)} (\mathbf{E}^{(1)})^4 \rangle. \tag{A.9}$$

The first term on the right-hand side of this relation, which is equal to $\langle \mathbf{j}^{(1)} \mathbf{E}^{(3)} \rangle$, vanishes and, hence, relation (A.9) leads to the well-known expression for $\chi_e^{(3)}$ [3, 4] (see formula (18)).

In the next approximation, we have

$$\langle \mathbf{j}^{(5)} \cdot \mathbf{E}^{(1)} \rangle = \langle \mathbf{j}^{(5)} \rangle \langle \mathbf{E}^{(1)} \rangle, \quad \langle \mathbf{j}^{(1)} \cdot \mathbf{E}^{(5)} \rangle = 0. \tag{A.10}$$

Using relation $\langle \mathbf{j}^{(5)} \rangle = \chi_e^{(5)} (\langle \mathbf{E} \rangle)^5$ and the definition of the quantity $\mathbf{j}^{(5)}(\mathbf{r})$ from Eqs. (A.3), from the first equality in (A.10) we obtain

$$\begin{aligned} \chi_e^{(5)} (\langle \mathbf{E} \rangle)^6 &= \langle \sigma \mathbf{E}^{(1)} \cdot \mathbf{E}^{(5)} \rangle \\ + 3 \langle \chi^{(3)} (\mathbf{E}^{(1)})^2 (\mathbf{E}^{(1)} \cdot \mathbf{E}^{(3)}) \rangle &+ \langle \chi^{(5)} (\mathbf{E}^{(1)})^6 \rangle. \end{aligned} \tag{A.11}$$

The first term on the right-hand side of this relation, which is equal to $\langle \mathbf{j}^{(1)} \cdot \mathbf{E}^{(5)} \rangle$, vanishes and, hence, we obtain from (A.11)

$$\chi_e^{(5)} = \langle \chi^{(5)} \mathbf{e}^6 \rangle + 3 \frac{\langle \chi^{(3)} \mathbf{e}^2 (\mathbf{e} \cdot \mathbf{E}^{(3)}) \rangle}{|\langle \mathbf{E} \rangle|^3}, \tag{A.12}$$

where $\mathbf{e}(\mathbf{r}) = \mathbf{E}^{(1)}(\mathbf{r})/|\langle \mathbf{E} \rangle|$. Higher-order approximations in $\langle \mathbf{E} \rangle$ can be analyzed similarly.

APPENDIX B

Let the local conductivity σ in a heterogeneous sample of volume V experience fluctuations in time:

$$\sigma(t, \tau) = \sigma(\mathbf{r}) + \delta\sigma(t, \mathbf{r}), \quad \overline{\delta\sigma} = 0. \tag{B.1}$$

Here, $\sigma(\mathbf{r})$ is the average value of conductivity and the bar indicates averaging over time. In this case, the total current through the sample also fluctuates: $I(t) = \bar{I} + \delta I(t)$. Let us define the time correlation functions for currents as

$$K(\tau) = \{\overline{\delta I(t + \tau) \delta I(t)}\} (\bar{I})^{-2}, \tag{B.2}$$

where the bar indicates averaging over time t as in relation (B.1). For a given potential difference, the total current is proportional to the average conductivity of the sample σ_e so that we have, instead of relation (B.2),

$$K(\tau) = \{\overline{\delta\sigma_e(t + \tau) \delta\sigma_e(t)}\} \sigma_e^{-2}, \tag{B.3}$$

where $\delta\sigma_e(t)$ is the fluctuating component of the quantity $\sigma_e(t)$. While deriving relation (B.3), we presumed the low-frequency nature of conductivity fluctuations so that the quasi-static approach is applicable and the current $I(t)$ is connected with voltage U through the conventional Ohm's law. Besides, we assume that the sample volume V is large enough so that the average conductivity σ_e coincides with the effective conductivity defined for $V \longrightarrow \infty$.

We will find the quantity $\delta\sigma_e$ by varying relation (5):

$$\delta\sigma_e(t) = \langle \delta\sigma \mathbf{e}^2 \rangle = \frac{1}{V} \int \delta\sigma(t, \mathbf{r}) \mathbf{e}^2(\mathbf{r}) d\mathbf{r}. \tag{B.4}$$

Here, we have taken into account the fact that $\langle \delta\mathbf{E} \delta\mathbf{E} \rangle = \langle \mathbf{j} \delta\mathbf{E} \rangle = \langle \mathbf{j} \rangle \langle \delta\mathbf{E} \rangle = 0$ since $\langle \delta\mathbf{E} \rangle = 0$. Using relation (B.4), we obtain

$$\begin{aligned} \overline{\delta\sigma_e(t + \tau) \delta\sigma_e(t)} &= \frac{1}{V^2} \int d\mathbf{r} \mathbf{e}^2(\mathbf{r}) \\ \times \int d\mathbf{r}' \mathbf{e}^2(\mathbf{r}') &\overline{\delta\sigma(t + \tau, r) \delta\sigma(t, \mathbf{r}')}. \end{aligned} \tag{B.5}$$

Following [3, 4], we make the natural assumption concerning the short-range (delta-functional in the macroscopic description) nature of correlations of the quantities $\delta\sigma_e(t, \mathbf{r})$:

$$\overline{\delta\sigma(t + \tau, \mathbf{r}) \delta\sigma(t, \mathbf{r}')} = \lambda(\tau, \mathbf{r}) \delta(\mathbf{r} - \mathbf{r}'). \tag{B.6}$$

In this case, we obtain from relations (B.3), (B.5), and (B.6)

$$K(\tau) = \frac{1}{V} \langle \lambda \mathbf{e}^4 \rangle \sigma_e^{-2}. \tag{B.7}$$

Passing to Fourier components in this relation, we arrive at formula (21). It should be noted that the quantity $K(\tau)$ is inversely proportional to volume V.

REFERENCES

1. A. M. Dykhne, Zh. Éksp. Teor. Fiz. **59**, 110 (1970) [Sov. Phys. JETP **32**, 63 (1970)].
2. B. Ya. Balagurov, Zh. Éksp. Teor. Fiz. **93**, 1888 (1987) [Sov. Phys. JETP **66**, 1079 (1987)].
3. D. Stround and P. M. Hui, Phys. Rev. B **37**, 8719 (1988).
4. D. J. Bergman, Phys. Rev. B **39**, 4598 (1989).
5. B. Ya. Balagurov and V. A. Kashin, Zh. Éksp. Teor. Fiz. **110**, 1001 (1996) [JETP **83**, 553 (1996)].
6. B. Ya. Balagurov, Zh. Éksp. Teor. Fiz. **88**, 1664 (1985) [Sov. Phys. JETP **61**, 991 (1985)].
7. B. Ya. Balagurov, Fiz. Tekh. Poluprovodn. (Leningrad) **20**, 1276 (1986) [Sov. Phys. Semicond. **20**, 805 (1986)].
8. A. L. Efros and B. I. Shklovskii, Phys. Status Solidi B **76**, 475 (1976).
9. D. B. Gingold and C. J. Lobb, Phys. Rev. B **42**, 8220 (1990).
10. O. Levy and D. J. Bergman, Phys. Rev. B **50**, 3652 (1994).
11. D. J. Frank and C. J. Lobb, Phys. Rev. B **37**, 302 (1988).
12. A. M. Satanin, S. V. Khor'kov, and A. Yu. Ugol'nikov, Pis'ma Zh. Éksp. Teor. Fiz. **62**, 301 (1995).
13. B. Ya. Balagurov, Zh. Éksp. Teor. Fiz. **82**, 1333 (1982) [Sov. Phys. JETP **55**, 774 (1982)].
14. B. Ya. Balagurov, Zh. Éksp. Teor. Fiz. **85**, 568 (1983) [Sov. Phys. JETP **58**, 331 (1983)].
15. B. Ya. Balagurov, Zh. Éksp. Teor. Fiz. **108**, 2202 (1995) [JETP **81**, 1200 (1995)].

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 93, No. 4, 2001, pp. 833–845.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 954–969.

SOLIDS
Electronic Properties

Magnetoresistance of an Asymmetric Quantum-Size Structure in a Parallel Magnetic Field: Field Asymmetry Independent of the Current Direction

A. A. Gorbatsevich[a], V. V. Kapaev[b], Yu. V. Kopaev[b], I. V. Kucherenko[b], O. E. Omel'yanovskiĭ[b, c], and V. I. Tsebro[b, c, *]

[a] *Moscow Institute of Electronic Engineering (Technological University), Moscow, 103498 Russia*
[b] *Lebedev Physical Institute, Russian Academy of Sciences, Leninskiĭ pr. 53, Moscow, 119991 Russia*
[c] *International Laboratory of Strong Magnetic Fields and Low Temperatures, 53421 Wroclaw, Poland*
**e-mail: tsebro@sci.lebedev.ru*

Received April 27, 2001

Abstract—A new phenomenon, viz., field-asymmetric transverse magnetoresistance of a doped asymmetric quantum-size structure discovered in a magnetic field parallel to the heteroboundary planes, is studied experimentally and theoretically. The magnetoresistance asymmetry relative to the field direction, which is independent of the direction of transport current, is observed when a lateral electric field is embedded in the structure with the help of alloyed metallic contacts. In the theoretical part of the paper, it is shown that the contribution to current, which is asymmetric in the magnetic field, can be consistently described in the framework of the theory of spontaneous current states and photovoltaic effect in systems without an inversion center; the reason behind the emergence of this current is associated with the asymmetry of the energy spectrum of charge carriers relative to the quasimomentum. It is shown that the change in the size and shape of Fermi contours in a magnetic field determines the magnitude of the strong negative magnetoresistance associated with the intersubband scattering under investigation and is found to be responsible for the emergence of a qualitatively new effect mentioned in the title of this paper. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The electronic properties of a low-dimensional systems are of considerable interest for fundamental science as well as for practical applications. Of special importance are investigations of 2D electron gas in a magnetic field, resulting in the discovery of integer [1] and fractional [2] quantum Hall effects. A 2D gas is realized in practice either at the heteroboundary between two semiconductors or in a quantum well. In both cases, the wave function of charge carriers is extended in a direction perpendicular to the heteroboundary plane and can vary in this direction under the effect of external factors (applied magnetic or electric field). Such a variation is manifested most clearly upon a transition to multilayered tunnel-coupled 2D systems asymmetric along the normal to the plane of heteroboundaries. If the role of external factor is played, for example, by a magnetic field, the abovementioned change in the wave function may noticeably affect the behavior of the magnetoresistance. This was demonstrated experimentally in [3–5], where the magnetoresistance of two tunnel-connected quantum wells was studied in a transverse (relative to the heterostructure plane) magnetic field.

A more significant change in the configuration of the wave functions of electrons in a direction perpendicular to heteroboundaries takes place when the magnetic field is applied along the plane of a heterostructure. In this case, the confining potential of the quantum well is supplemented by a magnetic potential of the oscillator type, which depends on the location of the center of the electron orbit in the magnetic field [6]. The magnetoresistance of nanostructures consisting of tunnel-connected quantum wells in such a geometry was investigated in [7–12]. Among other things, it was found that the magnitude of magnetoresistance is determined to a considerable extent by the change in the nature of intersubband scattering of charge carriers. The intersubband scattering, in turn, is determined by the magnitude of transferred momentum **Q**, which in the zeroth approximation is equal to the difference in the Fermi momenta in the subbands. The interband scattering probability is proportional to 1/**Q**. Upon a change in the magnetic field, the variations of Fermi momenta in different subbands differ considerably, which leads to the emergence of clearly manifested magnetoresistance. Since the relative change in the Fermi momenta in the subbands can be of either sign (depending on the system geometry and charge carrier concentrations in subbands), the magnetoresistance can be positive, or negative, or even alternating, as is indeed observed in experiments [8, 10, 12, 13].

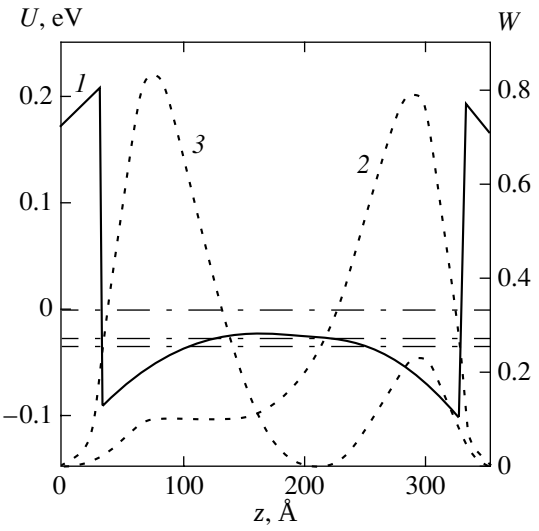

Fig. 1. Potential profile (curve *1*), position of size-quantization levels (dot-and-dash lines), and the probability distribution for the first two levels (curves *2*, *3*) for the structure under investigation in zero magnetic field.

It was found [7] that the magnitude of magnetoresistance in a parallel magnetic field depends on the direction of the magnetic field and the current; i.e., the magnetoresistance contains a term which is linear in the magnetic field parallel to heteroboundaries (and perpendicular to current) as well as in the electric field which is parallel to the current. The observed effect can be interpreted as a consequence of the electron density redistribution under the action of the Lorentz force. In the case when the scattering of charge carriers is different in different regions of a heterostructure, which is virtually always the case in heterostructures, the mobility of charge carriers changes upon a change in the direction of the current or the magnetic field, leading to the observed change in the resistance. The corresponding contribution to the current can be presented in the form

$$\delta\mathbf{j} = \alpha([\mathbf{P}\times\mathbf{H}]\cdot\mathbf{E})\mathbf{E}, \qquad (1)$$

where **H** is the magnetic field parallel to the layers, **E** is the electric field strength, and **P** is a certain polar vector perpendicular to the plane of the structure, which exists due to physical inequivalence of the opposite regions of the heterostructure, i.e., its asymmetry along the growth axis. It should be noted that according to the results of the experiments [7] and the corresponding theory [11], the effect associated with this mechanism is very weak.

In the present work, we describe a basically new effect [14], viz., magnetoresistance asymmetric to the magnetic field parallel to the layers of the structure, which is independent of the direction of the current. The experimental results obtained by us imply that a considerable asymmetry in magnetoresistance relative to the change in the magnetic field direction is observed in samples of a nanostructure with alloyed metallic contacts near which a depleted region with an embedded lateral electric field is formed (see Subsections 3.3.1 and 3.3.2). In this case, the magnetoresistance is independent of the direction of the current through the sample and, hence, the observed effect must correspond to a contribution to current in the form

$$\delta\mathbf{j} = \beta([\mathbf{P}\times\mathbf{H}]\cdot\mathbf{E}_0)\mathbf{E}, \qquad (2)$$

where $\mathbf{E}_0$ is a certain polar vector parallel to current, which is determined by the electric field embedded in the contact region. It will be shown in the theoretical part of this paper (Section 4) that the contribution of type (2) to the current can be described consistently using the theories of spontaneous current states [15] and the photovoltaic effect [16] in systems without an inversion center. The physical reason behind the emergence of the anomalous contribution to current is associated with the asymmetry of the energy spectrum of the investigated nanostructure relative to quasimomentum. Thus, while the change in the size of Fermi contours determines the value of magnetoresistance, which is associated with the intersubband scattering discussed by us here (Subsection 3.3.3), the change in their shape is responsible for the emergence of a qualitatively new effect defined by formula (2).

2. INVESTIGATED NANOSTRUCTURE AND ITS CHARACTERISTICS

The nanostructure investigated by us is a single undoped GaAs quantum well of width 300 Å, bounded on both sides by barrier layers of $Al_{0.34}Ga_{0.66}As$ (270 Å) uniformly doped with silicon up to the concentration $\sim 2\times 10^{18}$ cm^{-3}. The barrier layers are separated from the quantum well by undoped $Al_{0.34}Ga_{0.66}As$ spacers of width 100 Å. The entire structure is separated from the substrate by a thick (~0.5 μm) GaAs buffer layer and is covered by a protecting GaAs layer of thickness 100 Å. The profile of the bottom of the conduction band, the electron spectrum, and the wave function distribution were determined from the self-consistent solution to the system of Poisson and Schrödinger equations. It was found that the three subbands (E_1, E_2, and E_3) lying under the Fermi level are characterized by the Fermi energies $E_{F_1}\approx 32$ meV, $E_{F_2}\approx 26$ meV, and $E_{F_3} = 1$–2 meV. The distribution of the charge carrier concentration was calculated from the solution of the quantum-mechanical problem (using wave functions) in the region of the quantum well and according to classical formulas outside this region. Figure 1 shows the profile of the edge of the conduction band in the quantum well region, the position of the size-quantization levels (relative to the Fermi level), and the distribution of wave functions for the first two subbands. The distribution of the doping impurity in the system is such that the potential asymmetry in the quantum well is quite small: the potential difference ΔU at the right and left boundaries of the well amounts to only 12 meV. Nevertheless, the distri-

bution of the wave functions for the first subbands is characterized by a clearly manifested asymmetry (see Fig. 1).

3. MAGNETOTRANSPORT PROPERTIES (EXPERIMENTAL RESULTS AND DISCUSSION)

The magnetotransport properties of the above nanostructure were investigated using the standard dc method in the temperature range 2–300 K for various orientations of the magnetic field relative to the plane of the heterostructure and to the direction of transport current. Most measurements were made in the magnetic field region up to 75 kOe in a setup containing superconducting solenoid made of niobium–titanium cable, which made it possible to change the magnetic field direction during a run of measurements of field dependences of magnetoresistance. Separate measurements were made in strong magnetic fields up to 140 kOe created by a superconducting solenoid made of niobium–tin ribbon. Since the field reversal in such a solenoid is possible only up to a low backward field, the measurements in this case were made for a given orientation of the sample for a single chosen direction of the magnetic field.

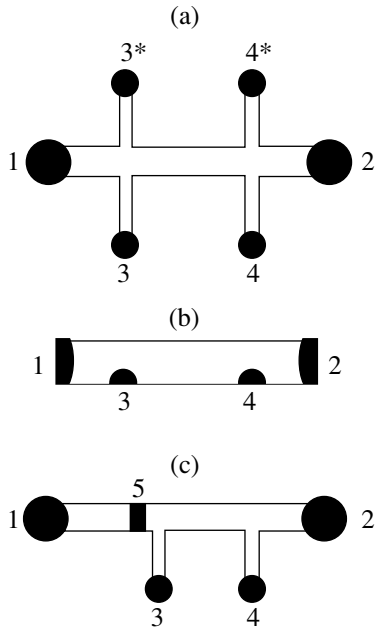

Fig. 2. Sample geometry and the shape of potential contacts: (a) lithographic potential contacts with the Hall geometry, (2) alloyed potential contacts, (c) combined contacts (*1* and *2* are current contacts and *3*–*5* are potential contacts).

3.1. Shape of Samples and Potential Contacts

The samples under investigation had different geometries and shapes of potential contacts (Fig. 2) depending on the type of the given experiment. The following three versions were used:

(1) lithographically prepared symmetric potential contacts of the Hall geometry (Fig. 2a);

(2) spaced alloyed metallic (indium) contacts (Fig. 2b);

(3) combined (one alloyed and several lithographic) contacts (Fig. 2c).

The geometry of the first version was used for characteristic measurements of the Shubnikov–de Haas (Sh–dH) effect and the quantum Hall effect (QHE). In the second and third versions with alloyed potential contacts, the contribution to the potential difference being measured comes from the contact regions of the nanostructure with embedded lateral electric fields; the asymmetry of the magnetoresistance relative to the field is observed precisely in this geometry. In the case when potential contacts are spaced to a certain distance (Fig. 2b), asymmetry is a difference effect which can be detected due to a difference in the values of the opposite embedded electric fields in the contact regions near the first and second potential contacts (see Subsection 3.3.1). In the third version (Fig. 2c), only one alloyed metallic potential contact was prepared. The second potential contact was a simple lithographic contact separated by a small distance from the alloyed contact in order to ensure a high enough value of the potential difference being measured, on the one hand, and to increase considerably the contribution to this potential difference from the contact region with an embedded electric field, on the other hand. In this case, the asymmetry of the magnetoresistance was noticeably enhanced (see Subsection 3.3.2).

3.2. Shubnikov–de Haas and Quantum Hall Effects

The Sh–dH effect and the QHE were measured in the standard configuration (the magnetic field is directed along the normal to the plane of the nanostructure, and the samples have the Hall geometry illustrated in Fig. 2a) primarily to characterize the nanostructure, i.e., to determine experimentally the parameters of its electron spectrum and to compare them with the calculated values. Figure 3a shows the results of measurements of the field dependences of magnetoresistance $R_{xx}(H)$ and the Hall resistance $R_{xy}(H)$ at $T = 4.2$ K. The Fourier analysis of the oscillating component of magnetoresistance gives two clearly manifested peaks corresponding to values 18.3×10^4 and 1.25×10^4. The values of charge carrier concentrations and the Fermi energy calculated by the formulas

$$n = \frac{e}{\pi c}\frac{1}{\Delta(1/H)}$$

and

$$E_F = \frac{e}{m^* c}\frac{1}{\Delta(1/H)}$$

(where $\Delta(1/H)$ is the period of oscillations in the reciprocal magnetic field), and corresponding to these two

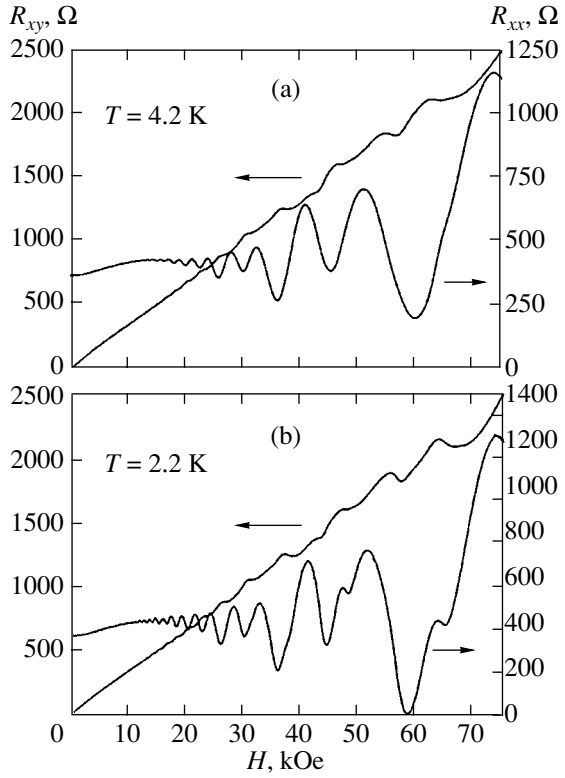

Fig. 3. Shubnikov–de Haas oscillations ($R_{xx}(H)$ curves) and field dependences of the Hall resistance ($R_{xy}(H)$ curves) at two temperatures: (a) 4.2 K and (b) 2.2 K.

values are $n = 8.8 \times 10^{11}$, 0.6×10^{11} cm^{-2}, and $E_F = 32$, 2 meV (for $m^* = 0.067m_0$). These values are close to the those calculated for subbands E_1 and E_3. Consequently, the contribution from subband E_2 to the Sh–dH oscillations is not manifested in the experiments probably due to the fact that the values of the Fermi energy and carrier concentrations in the first two subbands are close and the resolution is insufficient for a not very high mobility of carriers (see below) and for a high temperature and, second, due to the fact that the mobility of carriers in subband E_2 is lower than in subband E_1.[1]

The total carrier concentration n_H obtained from Hall measurements in weak magnetic fields ($H < 10$ kOe) is equal to 1.8×10^{12} cm^{-2}, while the measured value of electron mobility at T = 4.2 K is equal to 26000 cm^2/V s. A comparison of the Hall concentration n_H with the concentrations n_1 and n_3 obtained from the Sh–dH measurements shows that the values of the concentration in subbands E_1 and E_2 are indeed close, but the concentration in subband E_2 cannot be determined accurately since the Hall emf is determined by carriers in all the three subbands whose mobilities may differ considerably. It should be noted that the amplitude modulation of magnetoresistance oscillations in strong magnetic fields (Fig. 3) is probably associated with intersubband scattering whose amplitude is quite large (approximately equal to half the total scattering amplitude for charge carriers).

The field dependences of the Hall resistance $R_{xy}(H)$ (Fig. 3a) for $H > 12$ kOe display clearly manifested steps corresponding to the positions of minima of resistance R_{xx}. The values of R_{xy} at the steps coincides to within (3–4)% with the corresponding theoretical values of $\rho_{xy} = h/2eim$, where i is the number of filled Landau levels taking into account the spin and m is the number of identical electron layers connected in parallel ($m = 2$ in our case).

It can be seen from the results presented in Fig. 3b that the last two minima of the Sh–dH oscillations are shifted towards lower values of the magnetic field and split upon a decrease in temperature to 2.2 K. This is apparently associated with the spin splitting of Landau levels. It was shown in [17, 18] that the g-factor in a 2D gas at the AlGaAs/GaAs heteroboundary may attain high values ($g \sim 10$) for odd occupational numbers as a result of exchange interaction of electrons at the spin sublevels with quantum numbers i and $i - 1$.

It should be noted that the transverse magnetoresistance is positive in this configuration in the range of low magnetic fields (1.5–12 kOe). In very weak fields ($H <$ 1.5 kOe), a very weak (~0.5%) negative magnetoresistance is observed at T = 4.2 and 2.2 K, which is probably associated with the suppression of the interference correction to the metal-type conductivity [19].

3.3. Longitudinal and Transverse Magnetoresistance in a Parallel Magnetic Field

The dependence of the resistance of nanostructure samples on the strength of a magnetic field parallel to the plane of the 2D gas was investigated for two mutual orientations of the current vector **J** and the magnetic field vector **H**: **J** ⊥ **H** (transverse magnetoresistance) and **J** ∥ **H** (longitudinal magnetoresistance) in the temperature range 4.2–300 K. Measurements were made on samples with the three forms of potential contacts described above (see Fig. 2).

3.3.1. Magnetoresistance in the case of alloyed potential contacts with an embedded electric field. Figure 4 shows the magnetoresistance curves

$$\frac{\Delta R(H)}{R(0)} \equiv \frac{R(H) - R(0)}{R(0)}$$

measured in a wide temperature range (from 4.2 to 300 K) in the case of spaced alloyed potential contacts (Fig. 2b) for **J** ⊥ **H**. It can be seen that

(a) the magnetoresistance is negative in the entire temperature range;

[1] It can be seen from Fig. 1 that the wave function of electrons of subband E_1 has a clearly manifested peak at one of the heteroboundaries, while the wave function of subband E_2 has two peaks in the vicinity of all heteroboundaries. Consequently, the electron scattering probability at charge impurities located at the barriers in subband E_2 must be higher than in subband E_1.

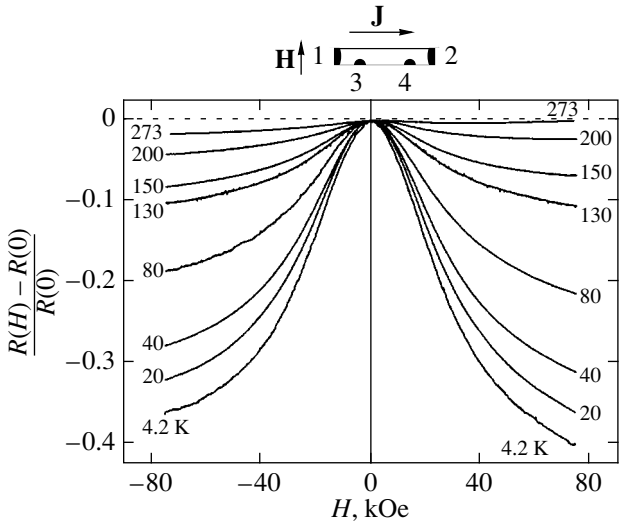

Fig. 4. Field dependences of magnetoresistance at different temperatures for **J** ⊥ **H**. The magnetic field vector is parallel to the plane of the 2D gas. The shape of the sample and potential contacts is shown at the top of the figure.

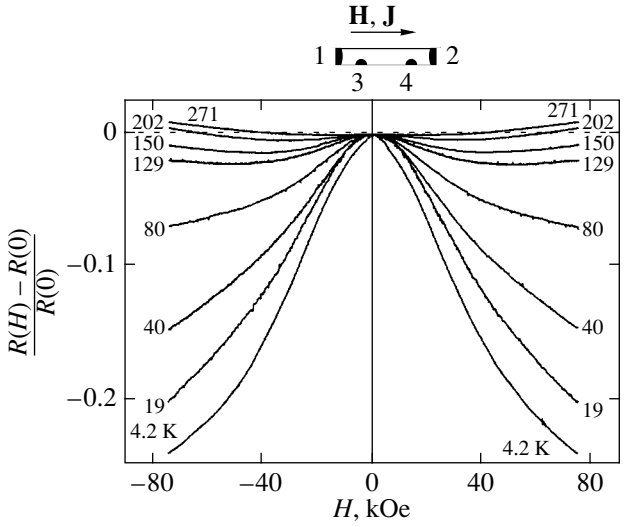

Fig. 5. Field dependences of magnetoresistance at different temperatures for **J** || **H**. The magnetic field vector is parallel to the plane of a 2D gas. The shape of the sample and potential contacts is shown at the top of the figure.

(b) the magnetoresistance strongly depends on temperature, increasing upon cooling and attaining a value of the order of –0.4 at T = 4.2 K and H = 75 kOe;

(c) the magnetoresistance is asymmetric in the magnetic field in the entire temperature range, the asymmetry changing its sign upon a change in temperature, passing through zero at a temperature of the order of 130 K.

It should be emphasized that the magnitude and shape of the normalized curves $\Delta R(H)/R(0)$ do not depend on the direction of the current through the sample and on its magnitude (at least, from 1 to 50 μA in the temperature range under investigation). This indicates that the Hall voltage emerging along the nanostructure axis for **J** ⊥ **H** does not affect the magnitude of magnetoresistance and its asymmetry.

In the case of the longitudinal magnetoresistance (**J** || **H**), the height and shape of the $\Delta R(H)/R(0)$ curves change, but these curves become completely symmetric (Fig. 5). The negative longitudinal magnetoresistance at T = 4.2 K and H = 75 kOe attains a value of about –0.25, which is considerably smaller than the transverse magnetoresistance in the given field at the given temperature (see Fig. 4).

Asymmetry of transverse magnetoresistance in a parallel magnetic field. Figure 6 presents the results on the asymmetry of the transverse magnetic resistance at two extreme temperatures, 4.2 and 273 K. These results were obtained on rectangular samples of the nanostructure having a size of 2 × 8 mm, in which two indium potential contacts separated by a distance of approximately 6 mm were alloyed at a temperature of 350°C. It can be clearly seen that the transverse magnetoresistance has different values for opposite directions of the magnetic field. For example, the asymmetry in the magnetoresistance relative to the magnetic field direction at T = 4.2 K attains ~7% in the field 75 kOe (see curves *1*). At T = 273 K, the magnitude of magnetoresistance itself is very small (~1%), but its asymmetry is manifested much more clearly. It can be seen from Fig. 6 (curves *2*) that one branch of the field dependence of magnetoresistance becomes nonmonotonic at high

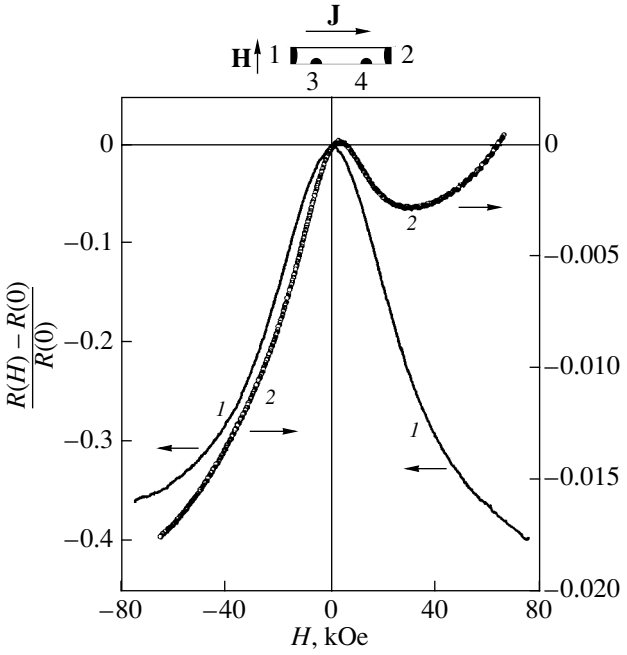

Fig. 6. Asymmetry of the field dependence of the transverse magnetoresistance at T = 4.2 K (curve *1*) and 273 K (curve *2*).

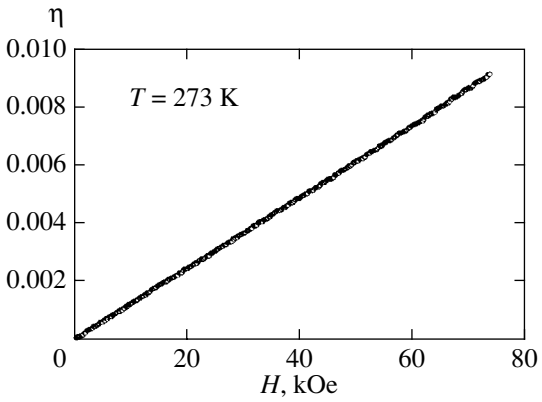

Fig. 7. Field dependence of asymmetry η of transverse magnetoresistance at 273 K.

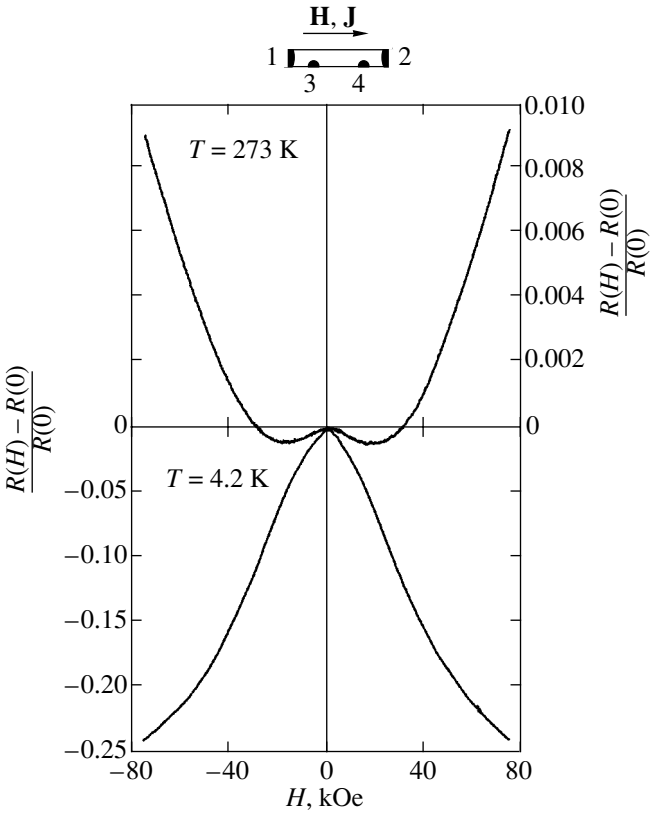

Fig. 8. Symmetric field dependences of the longitudinal magnetoresistance at different temperatures.

temperatures in contrast to the other branch, and the magnetoresistance even changes its sign in the vicinity of 30 kOe.

It was mentioned above that the magnitude and sign of transverse magnetoresistance asymmetry are independent of the direction of current for a given value of the magnetic field. If we define asymmetry $\eta(H)$ as the difference in the field dependences $\Delta R(H)/R(0)$ for opposite directions of the magnetic field, it is significant that the $\eta(H)$ dependence is a linear function of H in all the cases investigated by us and at all temperatures. This fact is illustrated in Fig. 7, where we deliberately choose the case when the field dependence $\Delta R(H)/R(0)$ is clearly nonmonotonic (magnetoresistance at $T = 273$ K, see curves *2* in Fig. 6). This leads to the conclusion that the asymmetric corntribution to conductivity (or current) is a linear function of the magnetic field.

If the sample is rotated so that the magnetic field vector becomes strictly parallel to the vector of current through the sample (longitudinal magnetoresistance), the magnitude and even the sign (at high temperatures in fields above 30 kOe) of the magnetoresistance change; it is significant, however, that the asymmetry of the $\Delta R(H)/R(0)$ curves disappears in this case (see Fig. 8).

3.3.2. Magnetoresistance in the case of combined potential contacts. Since in the case of spaced alloyed potential contacts the asymmetry of transverse magnetoresistance is a difference effect observed due to the difference in the magnitude of opposite embedded electric fields E_0 (see formula (2)) in the contact regions of the first and second contacts, it is interesting to measure the asymmetry of the magnetoresistance on samples with only one alloyed potential contact (see Fig. 2c). The other (lithographic) potential contact was arranged at a small (but not limiting) distance from the alloyed contact in order to ensure a sufficiently high level of the potential difference being measured, on the one hand, and to considerably increase the contribution to the total magnetoresistance from the depleted contact region of the alloyed potential contact with the embedded electric field, on the other hand.[2]

The results of such an experiment are presented in Fig. 9. Three potential contacts (depicted in the upper part of the figure) were used: an alloyed indium contact 5 and two lateral lithographic contacts 3 and 4 separated by quite different distances from contact 5. It can be seen that when the second lithographic potential contact is at a small distance from the alloyed contact ($d_{3-5} \approx 0.3$ mm in our case), the asymmetry of the magnetoresistance curves becomes more pronounced: it attains a value of the order of 50% at $T = 4.2$ K, while at $T = 273$ K, the magnetoresistance even changes its sign (it is positive for a certain direction of the field and negative for the opposite direction). As expected, the asymmetry of the magnetoresistance curves is weak when the second lithographic potential contact is far from the alloyed contact. In our case, when potential

[2] Alloying of a metallic (in our case, indium) contact to a certain depth is accompanied by the displacement of the potential barrier to the quantum well; as a result, a region depleted in charge carriers is formed under the contact and in its immediate vicinity. This depleted contact region possesses a very high resistance; for this reason, its (asymmetric relative to magnetic field) contribution to the magnetic resistance is considerable despite its small size (~1–10 μm) and determines the asymmetry of the transverse magnetoresistance of the entire sample.

contacts 4 and 5 are used (d_{4-5} = 6 mm), the asymmetry of the magnetoresistance curves does not exceed ~2% in compliance with the ratio of the distances d_{3-5} and d_{4-5}. Finally, in the case of the lithographic potential contacts 3 and 4, the magnetoresistance curves are completely symmetric (see curves *3*, *3*' in Fig. 9).

Thus, the experiments made on samples with an alloyed potential contact show that the effect of the asymmetry of the transverse magnetoresistance can be noticeably enhanced to the extent of contribution to the resistance being measured from the region of the nanostructure with an embedded lateral electric field, which is depleted in charge carriers.

3.3.3. Magnetoresistance and intersubband scattering. In this subsection, we will consider and analyze the problems and compare experimental and theoretical data associated with the negative magnetoresistance in a magnetic field parallel to the plane of the nanostructure and its possible qualitative explanation on the basis of the mechanism of intersubband scattering of charge carriers. It was mentioned in the Introduction that this mechanism is associated with the change in the size and shape of Fermi contours upon the application of an external magnetic field.

The high value of the negative magnetoresistance at low temperatures (in our case, $\Delta R(H)/R(0) \approx -0.4$ for $\mathbf{J} \perp \mathbf{H}$ and $\Delta R(H)/R(0) \approx -0.25$ for $\mathbf{J} \parallel \mathbf{H}$ in a field 75 kOe) is apparently a consequence of the suppression of the interband scattering in a magnetic field, which, like the intraband scattering at an ionized impurity, is determined by the electron mobility at low temperatures. Earlier, this mechanism of magnetoresistance at T = 4.2 K and $\mathbf{J} \perp \mathbf{H}$ was considered in [9, 10, 12, 20, 21] as applied to GaAs/AlGaAs heterojunctions and two-well structures.

It is well known (see, for example, [6]) that the electron states are subjected to a considerable modification in a parallel magnetic field, which leads to a displacement of the minima of the subbands to the k-space; the displacement of the minima of the corresponding subbands increases with the subband number and may attain a value of several millielectronvolts. Thus, the population of a certain subband in a magnetic field can be changed down to its complete depletion. Figure 10 shows the dispersion curves $E_n(k_x)$ and the corresponding Fermi contours in magnetic fields of 0, 40, and 80 kOe (the x and y axes lie in the plane of the nanostructure; the magnetic field is applied along the y axis), which were obtained for our nanostructure as a result of self-consistent solution of the system of Schrödinger and Poisson equations in a magnetic field. Obviously, the intersubband scattering in our case is determined mainly by E_1 and E_2 subbands with the smallest difference Δk_F in the wave vectors at the Fermi level. The population of the third subband is extremely low since the Fermi level lies near its bottom, and it is completely depleted in strong magnetic fields.

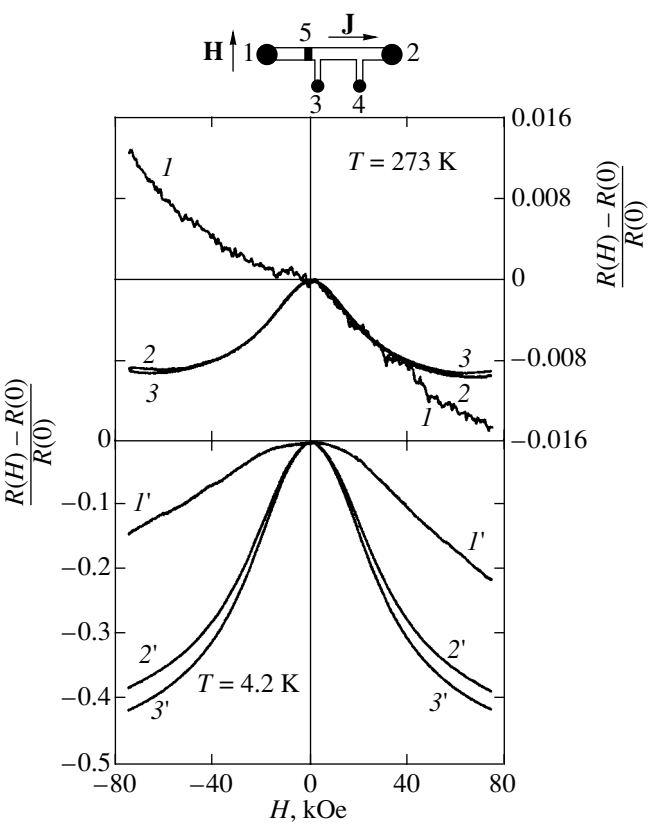

Fig. 9. Field dependences of magnetoresistance at 4.2 and 273 K in the case of combined potential contacts. Curves *1* and *1*' were obtained by measuring the potential difference between contacts 3 and 5, curves *2* and *2*', between contacts 4 and 5, and curves *3* and *3*', between contacts 3 and 4. The shape of the sample and the numbers of potential contacts are shown at the top of the figure.

It can be seen from Fig. 10 that as the value of the magnetic field increases, the dispersion relation $E_n(k_x)$ experiences considerable deformation even in a weakly asymmetric system like the nanostructure under investigation: the effective mass m becomes noticeably anisotropic (the effective mass m_x in the x direction increases with the field, while the effective mass m_y changes only slightly in the y direction). The asymmetry of the dispersion curves is manifested clearly, $E_n(k_x) \neq E_n(-k_x)$.

A complete description of the field dependence of the conductivity tensor $\sigma_{ij}(H)$ must include the solution of the corresponding kinetic equation (see, for example, [11]) and is a complicated problem. The features of magnetoresistance can be determined qualitatively by analyzing the dependence of the deformation of Fermi contours on the magnetic field. The existence of two subbands E_1 and E_2 with close energy values is responsible for a significant contribution of the mechanism of interband scattering of electrons to the magnetoresistance. The anisotropy of Fermi contours which appears upon an increase in the magnetic field leads to anisot-

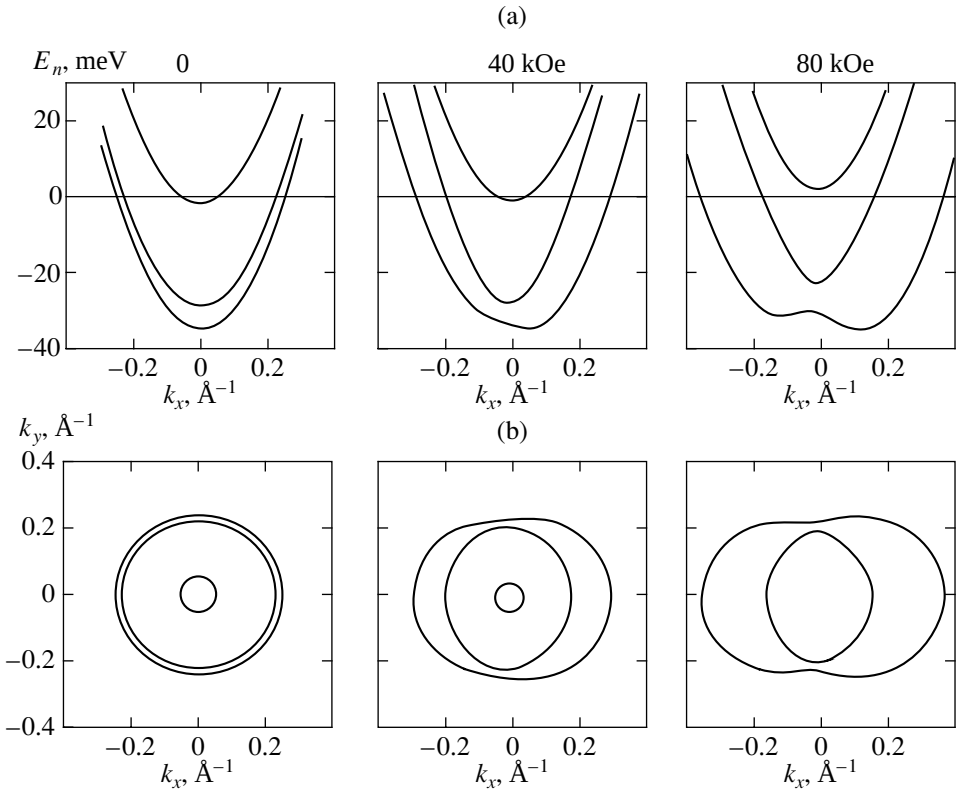

Fig. 10. Calculated dispersion curves (a) and Fermi contours (b) for different values of the magnetic field parallel to the plane of the nanostructure for the electron concentration $n = 2 \times 10^{12}$ cm^{-2}.

ropy of the characteristic time of intersubband relaxation and, hence, to a dependence of magnetoresistance on the mutual orientation of the current and the magnetic field. For anisotropic Fermi contours, conductivity σ_{xx} in the x direction is mainly determined by the states on the Fermi contours with small values of k_y, while conductivity σ_{yy} in the y direction is determined by states with small k_x. The intersubband scattering probability in this case depends on the separation Δk_F between the Fermi contours of the first two subbands in these directions. It can be seen from Fig. 10b that the values of Δk_F increase with the field in both directions. Since the scattering probability is inversely proportional to the value of momentum being transferred, the increase in the value of Δk_F with the magnetic field leads to a negative magnetoresistance, which is in accord with the experimental results. Calculations show that as the magnetic field ($\mathbf{H} \parallel y$) changes from zero to 100 kOe, the difference Δk_F in the y direction changes approximately by a factor of two, while the change in the transverse direction is almost tenfold. This explains the experimental fact of a strong change in the resistance with increasing field for the $\mathbf{J} \perp \mathbf{H}$ orientation. (It should be noted that the increase in the effective mass m_x is comparatively small and does not compensate the decrease in the scattering probability.)

An additional factor of the magnetoresistance anisotropy is that the overlapping of the wave functions for the first and second subbands for small k_x varies slowly with the magnetic field as compared to the case of small k_y. According to the results of calculations, the overlapping for small k_y first decreases upon an increase in H, attains its minimum value, and then increases, amplifying the scattering intensity.

The mobility of charge carriers at low temperatures is determined by the intraband scattering from a charged impurity and by the intersubband scattering. According to the results obtained in [20, 22], the probabilities of these processes are of the same order of magnitude. Consequently, the suppression of intersubband scattering in a magnetic field must considerably reduce the resistance, which is actually observed in our experiments. At high temperatures, the scattering at phonons becomes the dominating mechanism of intraband scattering. In this case, the decrease in the intersubband scattering probability in a magnetic field leads to a considerably smaller contribution to magnetoresistance.

3.3.4. Transverse magnetoresistance in strong magnetic fields. The measurements of transverse magnetoresistance in magnetic fields up to 140 kOe proved that the sign of the derivative of magnetoresistance is reversed at $H \approx 110$ kOe and the resistance of the nano-

structure starts increasing with the magnetic field. By way of an example, Fig. 11 shows a branch of the transverse magnetoresistance (*1*) and its derivative with respect to the field (*2*) at $T = 4.2$ K. The sign reversal of the magnetoresistance derivative, i.e., the transition to a positive magnetoresistance, indicates, as expected, the termination of the operation of the intersubband scattering mechanism in strong magnetic fields due to the withdrawal of the second subband E_2 beyond the Fermi level; as a result, only the E_1 subband remains below the Fermi level.

It is interesting to note that the curve describing the derivative of magnetoresistance clearly manifests singularities associated with quantization of the orbital motion of electrons in a magnetic field parallel to the plane of the nanostructure. Figure 11 shows, for example, that the curve of the magnetoresistance derivative displays typical kinks for the values of magnetic field for which the transverse size of the quantum well is such that it incorporates one ($n = 1$), one and a half ($n = 3/2$), and two ($n = 2$) electron orbits.[3] Thus, precision measurements of transverse magnetoresistance in a magnetic field parallel to the layers of a nanostructure make it possible to determine experimentally the transverse size of a quantum well to a high degree of accuracy and to compare it with the initial size at the growth stage of the nanostructure.

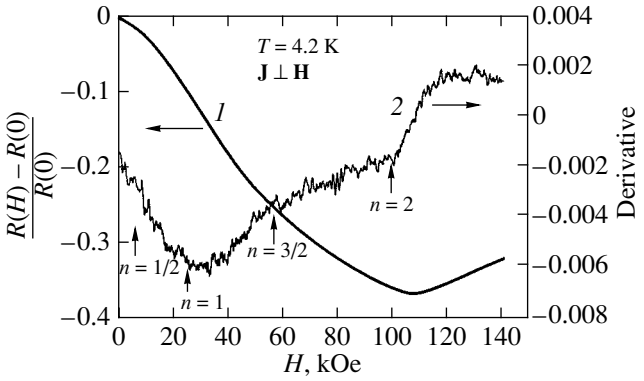

Fig. 11. One of the branches of transverse magnetoresistance at temperature 4.2 K (*1*) and its derivative (*2*) in magnetic fields up to 140 kOe. The sign reversal of the derivative of magnetoresistance takes place in the field 108 kOe. The arrows marking singularities (kinks) on the magnetoresistance derivative curve are indicated for the magnetic field values at which the transverse size of the quantum well corresponds to half ($n = 1/2$), one ($n = 1$), one and a half ($n = 3/2$), and two ($n = 2$) electron orbits for the orbital motion of electrons in a quantizing magnetic field parallel to the plane of the nanostructure.

4. PHENOMENOLOGICAL AND MICROSCOPIC DESCRIPTIONS OF MAGNETORESISTANCE ASYMMETRICIN IN THE MAGNETIC FIELD

A quantum-size structure in a longitudinal magnetic field is a system with a broken fundamental symmetry relative to inversion of coordinates and time [6]. The loss of the spacial inversion center is associated with a virtually inevitable asymmetry of the quantum well appearing due to inequivalence of the conditions of growth of the upper and lower heteroboundaries of the well. This asymmetry can be enhanced by forming multiwell structures containing unequivalent wells. It should be noted that there is formally no inversion center in any surface region due to the existence of the normal to the surface. The system under investigation can be characterized by the polar time-odd vector **T**:

$$\mathbf{T} \propto \mathbf{H} \times \mathbf{P}. \tag{3}$$

Here, **H** is the external magnetic field and **P** is a certain polar vector directed along the normal to the plane of the well. Vector **T** is similar in transformation properties to the quasimomentum vector **k**. Consequently, the product $\mathbf{T} \cdot \mathbf{k}$ is invariant and the energy spectrum, which may contain all possible invariants, is asymmetric relative to the wave vector:

$$\varepsilon(\mathbf{k}) \neq \varepsilon(-\mathbf{k}). \tag{4}$$

This asymmetry may lead to peculiar macroscopic properties of nanostructures. For example, according to theoretical publications [6, 23], anomalous photovoltaic and magnetoelectric effects emerge in such a system under nonequilibrium conditions. In order to interpret these and other effects qualitatively, it is expedient to use the concept of vector **T** as the density vector of the toroidal moment of the system (toroidal multipoles form the third independent family of multipoles in addition to the electric and magnetic moments [24]). Since the toroidal moment, momentum, and current possess the same symmetry, the existence of a macroscopic spontaneous current in the system, which is proportional to the density vector of the toroidal moment,

$$\mathbf{j} = \beta \mathbf{T} \tag{5}$$

does not contradict the requirements of symmetry.

However, this expression for the current density vector under equilibrium conditions is forbidden by the gauge invariance requirements [25]. Under nonequilibrium conditions, however, a current of the form (5) may exist [6]. In this case, parameter β in formula (5) is a dissipative constant. Since both vectors **j** and **T** change their sign upon time inversion, constant β must be proportional to an even power of relaxation time. The deviation from equilibrium can be due to photoexcitation [16]. In this case, Eq. (3) describes the anomalous photovoltaic effect (this effect is anomalous since it exists in the absence of the electrochemical potential gradient). This effect can be observed not only in systems

[3] In Fig. 11, the singularity of the magnetoresistance derivative for $n = 1/2$ is manifested very weakly. However, precision experiments in weak magnetic fields indicate that a clearly manifested jump in the second derivative of magnetoresistance takes place at this point also.

with an asymmetric spectrum, but also in systems without an inversion center [26]. The mechanisms behind the emergence of the effect in these systems are substantially different. In nonmagnetic noncentrosymmetric media, it is due to the emergence of the asymmetric component of the distribution function under an isotropic perturbation of the system. The description of the effect in this case falls outside the scope of the Born approximation in scattering [26]. On the contrary, the anomalous photovoltaic effect in systems with broken magnetic symmetry and with an asymmetric spectrum can be described even in the relaxation time approximation and is associated with the very fact that the system is not in equilibrium [16]. However, this effect emerges only if the nonequilibrium distribution function is essentially out of equilibrium and should not be reducible to a function depending only on the energy of particles [16]. Functions of this type appear in the theory if several relaxation channels "operate" simultaneously. As in conventional nonmagnetic media [26], nonequilibrium currents of the form (5) may also appear for sources of deviation from equilibrium other than photoexcitation, e.g., nonequilibrium heating [27] (in the absence of equilibrium between the electron and phonon subsystems).

Let us consider the situation when the system is out of equilibrium due to a direct current passing through it. The general expression for the current density has the conventional form

$$\mathbf{j} = \hbar^{-1}\int\frac{\partial\varepsilon(\mathbf{k})}{\partial\mathbf{k}}f(\mathbf{k})\frac{d^D k}{(2\pi)^D}, \tag{6}$$

where $\varepsilon(\mathbf{k})$ is the energy spectrum, $f(\mathbf{k})$ is the charge carrier distribution function, and D is the dimensionality of the $\mathbf{k}$ space. The distribution function $f(\mathbf{k})$ can be determined from the solution of the kinetic equation

$$\left(\mathbf{v}\frac{\partial}{\partial\mathbf{r}} - \frac{e}{\hbar}\left(\mathbf{E} + \frac{1}{c}\mathbf{v}\times\mathbf{H}\right)\frac{\partial}{\partial\mathbf{k}}\right)f(\mathbf{k},\mathbf{r}) = I_{\text{sct}}[f], \tag{7}$$

where $\mathbf{E}$ is the total electric field acting on charge carriers, including the external electric field associated with the potential applied to the structure and the internal embedded field of the contact potential and/or of the bulk charge, $\mathbf{v} = (1/\hbar)\partial\varepsilon(\mathbf{k})/\partial\mathbf{k}$ is the velocity of charge carriers, and $I_{\text{sct}}[f]$ is the scattering integral.

The energy spectrum $\varepsilon(\mathbf{k})$ of charge carriers appearing in formulas (6) and (7) must be determined from the solution of the Schrödinger equation for a quantum structure in a magnetic field. We choose the normal to the plane of the quantum well (size-quantization axis) along the z axis. We assume that the magnetic field is directed along the y axis and the electric current, along the x axis. We also assume that an embedded electric field associated with the structural inhomogeneity of the material (the field of the contact potential or the field of the space charge) is directed along the current. We denote by $\varphi(x)$ the total electrostatic potential of the embedded and external fields and assume that it is smooth and satisfies the semiclassical approximation. The potential $U(z)$ of the quantum structure will be described by using the envelope function method (effective mass method). The wave function of charge carriers in this case is sought in the form

$$\psi_{n,k_x,k_y}(\mathbf{r}) = \exp\{i(k_x x + k_y y)\}\phi_{n,k_x}(z).$$

The expression for the energy of charge carriers has the form

$$\begin{aligned}\varepsilon(k_x, k_y, x) &= \varepsilon_n(k_x, k_y) - e\varphi(x)\\ &= \frac{\hbar^2 k_y^2}{2m} + \varepsilon_n(k_x) - e\varphi(x),\end{aligned} \tag{8}$$

where $\varepsilon_n(k_x)$ is the quantized energy spectrum, which can be determined from the Schrödinger equation

$$\begin{aligned}\left[-\frac{\hbar^2}{2m}\nabla_z^2 + \frac{\hbar}{2}\omega_0\lambda^{-2}(z-z_0)^2 + U(z)\right]\phi_{n,k_x}(z)\\ = \varepsilon_n(k_x)\phi_{n,k_x}(z)\end{aligned} \tag{9}$$

(here, $\omega_0 = eH/mc$ is the cyclotron frequency, $\lambda = (\hbar c/eH)^{1/2}$ is the magnetic length, and $z_0 = k_x\lambda^2$ is the center of the orbit of the wave function for an electron in a magnetic field). In the absence of the quantum structure potential $U(z)$ (in a homogeneous medium), the energy of a particle in a magnetic field does not depend on the position of the center of the orbit and the energy levels ε_n are degenerate in k_x. In the presence of potential $U(z)$, the energy levels acquire a dispersion $\varepsilon_n = \varepsilon_n(k_x)$ due to the dependence of the particle energy on the position z_0 of the orbit center in a magnetic field. If potential $U(z)$ is spatially asymmetric, $U(z) \neq U(-z)$, the energy $\varepsilon(k_x)$ automatically becomes an asymmetric function of quasimomentum (see relation (4)) since the coordinates $z_0 = k_x$ and $-z_0 = -k_x$ of the orbit center are inequivalent.

The presence of the quantizing potential of the heterostructure considerably complicates the solution of the kinetic equation. The nonlocality of the scattering integral for the motion of charge carriers in a direction perpendicular to the magnetic field cannot be resolved analytically even in the case of a symmetric quantum well in a magnetic field parallel to the plane of the well, and the nonlocal renormalization of the velocity of carriers must be taken into account [28]. In this case, the kinetic equation can be solved only numerically even for the simplest model potential [29]. Asymmetry of the quantizing potential complicates the problem still further. In order to clarify the qualitative aspects of the phenomenon under investigation, we use the artificial relaxation time approximation for writing the scattering integral in the kinetic equation (7):

$$I_{\text{sct}}[f^{(i)}] = -\frac{f^{(i)}(\mathbf{k},\mathbf{r})}{\tau_i}, \tag{10}$$

where $i = 1, 2, \ldots$ is the order of expansion of the distribution function in the small parameter of the problem and τ_i is the corresponding relaxation time. In a spatially symmetric system, $i = 1$ corresponds to the distribution function component asymmetric in $\mathbf{k}$, while $i = 2$ corresponds to the nonequilibrium correction to the symmetric component of the distribution function. In an asymmetric system with the spectrum asymmetric in quasimomentum, the separation of the symmetric and asymmetric components is not expedient.

In the general case, the diffusion and the field terms on the left-hand side of Eq. (7) are separately not small even in the case of a slight deviation from equilibrium (weak current) in a system with an embedded potential. However, these terms exactly compensate each other in equilibrium. If the departure from equilibrium is insignificant, we can choose for the small parameter the density j of electric current passing through the system. The distribution function in this case can be presented in the form of a series,

$$f = f^{(0)}(1 + a_1(E_0)j + a_2(E_0)j^2 + \ldots),$$

where $f^{(0)}$ is the equilibrium distribution function and $a_i(E_0)$ are the expansion coefficients depending only on the magnitude E_0 of the embedded field; these coefficients can be calculated using perturbation theory assuming that the embedded field E_0 is weak. As a result, in the expansion for the distribution function, we have

$$\begin{aligned} f &= f^{(0)} + f^{(1)} + f^{(2)} + \ldots \\ &= f^{(0)} + f^{(1)} + f^{(21)} + f^{(22)} + \ldots \\ &= f^{(0)}(1 + a_1^0 j + a_1^1 E_0 j + a_2^0 j^2 + \ldots), \end{aligned}$$

where the second term ($f^{(1)}$) describes the conventional ohmic current in the linear approximation, the fourth term ($f^{(22)}$) describes the nonlinear correction to the current, while the third term ($f^{(21)}$) describes the correction to the ohmic current and to the linear conductivity, which is associated with the macroscopic inhomogeneity of the system. It will be shown below that this correction can be a nontrivial function of the magnetic field.

Using the kinetic equation (7), we obtain the following expression for the first-order (in current) contribution to the distribution function:

$$f^{(1)} = \frac{\tau_1}{\tilde{T}}\mathbf{v}\nabla F f^{(0)}, \tag{11}$$

where $\tilde{T}$ is the temperature and F is the Fermi quasilevel (electrochemical potential): $\nabla F = -e\nabla(\varphi - \varphi_0)$; here, φ and φ_0 are the potentials of the total and embedded electric fields, respectively: $\nabla\varphi = E$, $\nabla\varphi_0 = E_0$. In the equilibrium stare, the Fermi level $F_0 = \text{const}$.

The substitution of relation (11) into formula (6) for current leads to the conventional expression for ohmic current:

$$j_0 = \frac{ne\tau_1}{m}\nabla F = n\mu\nabla F, \tag{12}$$

where μ is the mobility and n is the charge carrier concentration.

Iterating Eq. (7) twice, we obtain the following expression for the second-order contribution $f^{(2)}$:

$$\begin{aligned} f^{(2)} = -\frac{e^2\tau_1\tau_2}{\hbar\tilde{T}}&\left(\left(\mathbf{E} + \frac{1}{c}\mathbf{v}\times\mathbf{H}\right)\frac{\partial}{\partial\mathbf{k}}\right) \\ &\times\left(\mathbf{v}\frac{\partial}{\partial\mathbf{r}}(\varphi(x) - \varphi_0(x))\right)f^{(0)}. \end{aligned} \tag{13}$$

This relation immediately implies that in the case of a charge carrier spectrum asymmetric in quasimomentum ($\mathbf{v}(\mathbf{k}) \neq -\mathbf{v}(-\mathbf{k})$), the substitution of expression (13) into relation (6) for current gives a term bilinear in the magnetic field and the electrochemical potential gradient, whose symmetry differs from the symmetry of the contribution to the current, associated with the Hall effect.

The analytic structure of the expression for current can be investigated by analyzing a model asymmetric dispersion relation of the form

$$\varepsilon(\mathbf{v}) = \frac{mv^2}{2}(1 + \boldsymbol{\Upsilon}\mathbf{v}), \tag{14}$$

where the velocity $\mathbf{v} = \hbar\mathbf{k}/m$ and $\boldsymbol{\Upsilon}$ is the asymmetry parameter of the spectrum (4) for charge carriers, which is proportional to the toroidal moment (3). The left-hand side of Eq. (9) and, hence, the energy spectrum are invariant to the simultaneous sign reversal of the magnetic field $\mathbf{H}$ and quasimomentum k_x. Thus, the spectrum is also asymmetric relative to the magnetic field, the asymmetry parameter $\boldsymbol{\Upsilon}$ of the spectrum being proportional to odd powers of the magnetic field,

$$\boldsymbol{\Upsilon}(\mathbf{H}) = -\boldsymbol{\Upsilon}(-\mathbf{H}), \tag{15}$$

in compliance with the general symmetry considerations (3).

Using expressions (14), (13), and formula (6) for current, we obtain the following expression for the contribution to current in the second order of perturbation theory:

$$\mathbf{j}_2 = \mu[(\mathbf{E}_0\cdot\boldsymbol{\Upsilon})\mathbf{j}_0 + (\mathbf{j}_0\cdot\boldsymbol{\Upsilon})\mathbf{E}_0], \tag{16}$$

where $\mathbf{j}_0$ is the ohmic current (12).

For the electric conductivity tensor σ_{ik} defined by

$$j_i = \sigma_{ik}(\nabla F)_k,$$

we obtain from relations (12) and (16) the following expression in the second order of perturbation theory:

$$\sigma_{ik} = n\mu^2(\mathbf{E}_0 \cdot \boldsymbol{\Upsilon}\delta_{ik} + E_{0i}\Upsilon_k). \quad (17)$$

For parallel $\mathbf{E}_0$ and $\mathbf{j}_0$, the first term on the right-hand side of Eq. (17) gives the contribution to the diagonal component to conductivity, which is odd in the magnetic field by virtue of relation (15). It follows from formula (16) that such a contribution to conductivity (resistance) exists only if the current flows through the region in which the embedded field E_0 differs from zero. The larger the value of the embedded field, the stronger (other conditions being equal) must be the manifestation of the magnetoresistance asymmetry relative to magnetic field, which agrees qualitatively with the main features of the asymmetric resistance (relative to the magnetic field) observed by us for the first time in [14] and described in detail in Subsections 3.3.1 and 3.3.2 of this paper. It should be noted that the magnetic field asymmetric contribution to magnetoresistance, which was considered above and is associated with the emergence of additional contribution (16) to the current, is a first-order effect relative to the external field (electrochemical potential gradient). For this reason, the corresponding linear responses (electrical conductivity and magnetoresistance) do not change upon the sign reversal of the current. On the contrary, the magnetoresistance asymmetry relative to the change in magnetic field and current, which is associated with the redistribution of charge and current density under the action of the Lorentz force [7], is a second-order effect in the external field, which determines its smallness.

Expression (16) for current suggests that a number of new effects may also exist. If the directions of the current and of the embedded field do not coincide (e.g., the magnetic field and the current are parallel, while the embedded electric field is perpendicular to them), the contribution to the resistance of the system described by the first term on the right-hand side of Eq. (16) is nevertheless an asymmetric function of the magnetic field direction and is independent of the direction of the current. The second term in Eq. (16) for parallel $\mathbf{E}_0$ and $\mathbf{j}_0$ is completely similar to the first term. If, however, the embedded field and the current are not parallel, this term describes, in the Hall geometry, the emergence of an emf in a direction parallel to the magnetic field provided that there exists an embedded electric field parallel to the magnetic field. It should be noted that the absence of a term proportional to γ itself (or to the toroidal moment vector $\mathbf{T}$ since $\gamma \propto \mathbf{T}$) in expression (16) for current is a consequence of simplifications (10) used in the solution of the kinetic equation. Such a term will be present in the expression for current if the nonlocal renormalization of velocity is taken into account [29].

5. CONCLUSIONS

In this work, we investigated the anomalous (in respect of asymmetry) magnetic field dependence of the resistance of a nonhomogeneous quantum-size heterostructure. In fact, we proved that macroscopically inhomogeneous heterostructures are nontrivial physical objects exhibiting new physical properties which can be interpreted from the position of macroscopic symmetry. In the theoretical part of the present paper (Section 4), we described a number of new galvanomagnetic phenomena which may exist in such systems. New effects appear as a result of simultaneous action on charge carriers by external electric and magnetic fields as well as the internal (embedded) field of the space charge or contact potential. In our experiments, like in most other electrophysical and galvanomagnetic experiments, we measured the global characteristics of the medium (such as the total resistance and conductance) rather than local characteristics. In this case, a situation can emerge, in principle, when the directions of the current and the magnetic field are not collinear in some regions in the system with a strictly parallel orientation of the external magnetic field and the current contacts due to a nonuniform distribution of the current. The macroscopic resistance measured in this case is determined by the longitudinal as well as the transverse (odd in magnetic field) resistances.

It should be noted that the magnetoresistance asymmetry relative to the magnetic field was observed earlier in mesoscopic conductors [30] and in faceted heterostructures in a nonuniform transverse magnetic field [31]. This effect in mesoscopic conductors was interpreted precisely as the result of contribution of the nondiagonal (Hall) component to the conductance of the system being measured [32, 33]. The symmetry of the system in the case of faceted heterostructures is similar in many respects to the symmetry of the experiment described by us here. The role of vector $\mathbf{T}$ in this case is played by the curl of the normal component of the magnetic field, which also leads to the asymmetry of the spectrum relative to quasimomentum [34]. It should be noted that, in accordance with the concepts developed in the present work as well as in [6, 23], the asymmetry of the energy spectrum relative to quasimomentum must lead, in faceted heterostructures in a tilted magnetic field, to the magnetoelectric and photovoltaic effects described earlier [6, 23] and to the asymmetry of the longitudinal (non-Hall) component of magnetoresistance described in this work. The authors of [31] interpret the experimental results by using, in fact, the approach developed for mesoscopic conductors and indicate the possible curvilinear geometry of the distribution of current, leading to the entanglement of the longitudinal and the transverse (Hall) components of magnetoresistance in the observed response as the mechanism responsible for the effect. It is significant, however, that in this case (in contrast to the case described by us here), the effect will depend on the

direction of the transport current. In order to determine the actual mechanism of the magnetoresistance asymmetry in faceted heterostructures in a tilted magnetic field, it would be expedient to carry out more detailed experiments on samples with different topologies of contacts, which will make it possible to compare different geometries of current spreading.

ACKNOWLEDGMENTS

This work was supported by the Interindustry Research and Development Program of the Russian Federation "Physics of Solid Nanostructures," Federal Scientific and Technical Program "Modern Trends in the Physics of Condensed Media" ("Surface Atomic Structures"), and by the Federal Target Program "Integration."

REFERENCES

1. K. von Klitzing, G. Dorda, and M. Pepper, Phys. Rev. Lett. **45**, 494 (1980).
2. D. C. Tsui, H. L. Stormer, and A. C. Gossard, Phys. Rev. Lett. **48**, 1559 (1982).
3. G. S. Boebinger, H. W. Jang, L. N. Pfieffer, and K. W. West, Phys. Rev. Lett. **64**, 1793 (1990).
4. A. G. Davies, C. H. W. Barnes, K. R. Zolleis, *et al.*, Phys. Rev. B **54**, R17331 (1996).
5. I. S. Millard, N. K. Patel, M. Y. Simmons, *et al.*, J. Phys.: Condens. Matter **8**, L311 (1996).
6. A. A. Gorbatsevich, V. V. Kapaev, and Yu. V. Kopaev, Pis'ma Zh. Éksp. Teor. Fiz. **57**, 565 (1993) [JETP Lett. **57**, 580 (1993)].
7. H. Sakaki, H. Ohno, S. Nishi, and J. Yoshino, Physica B (Amsterdam) **117–118**, 703 (1983).
8. D. R. Ledley, R. J. Nicholas, J. J. Harris, and C. T. Foxon, Semicond. Sci. Technol. **5**, 1081 (1990).
9. J. M. Heisz and E. Zaremba, Semicond. Sci. Technol. **8**, 575 (1993).
10. Y. Berk, A. Kamenev, A. Palevsky, *et al.*, Phys. Rev. B **51**, 2604 (1995).
11. J. M. Heisz and E. Zaremba, Phys. Rev. B **53**, 13594 (1996).
12. T. Jungwirth, T. S. Lay, L. Smrcka, and M. Shayegan, Phys. Rev. B **56**, 1029 (1997).
13. D. R. Ledley, R. J. Nicholas, J. J. Harris, and C. T. Foxon, in *Proceedings of the 20th International Conference on Physics of Semiconductors*, Ed. by Anastassakis and J. D. Jaonopopoulos (World Sci., Singapore, 1990), p. 1609.
14. A. A. Gorbatsevich, V. V. Kapaev, Yu. V. Kopaev, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **68**, 380 (1998) [JETP Lett. **68**, 404 (1998)].
15. A. A. Gorbatsevich and Yu. V. Kopaev, Ferroelectrics **161**, 321 (1994).
16. Yu. A. Artamonov, A. A. Gorbatsevich, and Yu. V. Kopaev, Zh. Éksp. Teor. Fiz. **101**, 557 (1992) [Sov. Phys. JETP **74**, 296 (1992)].
17. A. Usher, R. J. Nicholas, J. J. Harris, and C. J. Foxon, Phys. Rev. B **41**, 1129 (1990).
18. I. V. Kukushkin, Doctoral (Phys.–Math.) Dissertation (Chernogolovka, 1990), p. 142.
19. B. L. Altshuler and A. G. Aronov, *Electron–Electron Interactions in Disordered Systems*, Ed. by A. L. Efros and M. Pollak (North-Holland, Amsterdam, 1985).
20. S. Mori and T. Ando, J. Phys. Soc. Jpn. **48**, 865 (1980).
21. P. Streda, P. Vasek, and M. Curk, Phys. Rev. B **51**, 11144 (1995).
22. J. J. Harris, D. E. Lacklisson, C. T. Foxon, *et al.*, Semicond. Sci. Technol. **2**, 783 (1987).
23. A. A. Gorbatsevich, V. V. Kapaev, and Yu. V. Kopaev, Ferroelectrics **161**, 303 (1994).
24. V. M. Dubovik and L. A. Tosunyan, Fiz. Elem. Chastits At. Yadra **14**, 1193 (1983) [Sov. J. Part. Nucl. **14**, 504 (1983)].
25. A. A. Gorbatsevich, Zh. Éksp. Teor. Fiz. **95**, 1467 (1989) [Sov. Phys. JETP **68**, 847 (1989)].
26. V. I. Belinicher and B. I. Sturman, Usp. Fiz. Nauk **130**, 415 (1980) [Sov. Phys. Usp. **23**, 199 (1980)].
27. O. V. Kibis, Zh. Éksp. Teor. Fiz. **115**, 959 (1999) [JETP **88**, 527 (1999)].
28. H. Tang and P. N. Butcher, J. Phys. C **21**, 3313 (1988).
29. H. Tang and P. N. Butcher, J. Phys. C **21**, 3959 (1988).
30. R. A. Webb, S. Washburn, C. Umbach, and R. A. Laibowitz, Phys. Rev. Lett. **54**, 2696 (1985).
31. M. L. Leadbeater, C. L. Foden, J. H. Burroughes, *et al.*, Phys. Rev. B **52**, R8629 (1995).
32. M. Buttiker, Phys. Rev. Lett. **57**, 1761 (1986).
33. M. Ma and P. A. Lee, Phys. Rev. B **35**, 1448 (1987).
34. J. E. Muller, Phys. Rev. Lett. **68**, 385 (1992).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 93, No. 4, 2001, pp. 846–852.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 970–978.

SOLIDS
Electronic Properties

Rehybridization of the Atomic Orbitals and the Field Electron Emission from Nanostructured Carbon

A. N. Obraztsov[a,*], A. P. Volkov[a], A. I. Boronin[b], and S. V. Koshcheev[b]

[a]*Moscow State University, Moscow, 119899 Russia*
[b]*Boreskov Institute of Catalysis, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia*
**e-mail: obraz@acryst.phys.msu.su*

Received March 30, 2001

Abstract—The field electron emission, structural features, and electronic properties of carbon films obtained by chemical vapor deposition were experimentally studied. It is shown that the field electron emission from the films composed of spatially oriented carbon nanotubes and platelike graphite nanocrystals is observed for the electric field strength lower by one to two orders of magnitude as compared to the values characteristic of the metal emitters. Experimental data reported for the first time are indicative of a local decrease in the electron work function in such carbon film materials as compared to that in graphite. A model of the emission center is proposed and a mechanism of the field electron emission from nanostructured carbon is described. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The main advantages of field (or cold) cathodes in comparison to the incandescent electron emitters are the absence of energy losses for the electron excitation, the high density of the emission current, and the narrow distribution of the emitted electron beam with respect to the emission direction and particle velocity [1, 2]. At the same time, an extremely high electric field strength (about 100 V/μm) required in most cases for the field electron emission brings considerable problems in the cold cathode applications. In particular, such cathodes have to be manufactured in the form of thin points or blades (see, e.g., [3]) in order to provide for the necessary local field amplification.

An alternative approach consists in fabricating cathodes from semiconductor materials possessing low or negative electron affinity, which allows the field strength and, hence, the potential difference between cathode and anode to be reduced. As a result, problems related to the damaging effects of ions bombarding the cathode surface and of the high-strength electric field are eliminated [4, 5].

A typical example of the semiconductor material suited for these purposes is offered by diamond. Some faces of the diamond single crystals are characterized by a negative electron affinity [6], which provides conditions for the obtaining of cold emission at low threshold values of the field strength in various diamond or diamond-like materials [7]. Unfortunately, the concentration of free electrons in pure diamond is too low for the emission current to be sufficiently high. At the same time, the problem of doping diamonds so as to create shallow donor levels and increase the free electron concentration is still unsolved for various (including fundamental) reasons [8].

Alternatively, free electrons in the diamond structure can be generated under the action of UV or X-ray radiation, which is analogous to the excitation process taking place during the photoelectron spectroscopy measurements (by which, in particular, the negative electron affinity of diamond was found—see, e.g., [6]). Electrons can be also injected into the conduction band of diamond from some other material. The main factors determining the emission efficiency are related to the recombination processes and the transport of electrons to the cathode surface [6, 9].

An analysis of numerous experimental data reported so far shows that the efficiency of cold emission in diamond materials increases with the content of non-diamond carbon present in the form of extended structural defects of various origin [10–12] or graphite-like carbon inclusions formed at the grain boundaries in polycrystalline films obtained by chemical vapor deposition (CVD) [13, 14] and in graphite–diamond compounds [15, 16].

Our recent investigations [17–19] of CVD diamond films showed for the first time that gradual variation in the ratio of the diamond and graphite-like components is accompanied by a monotonic decrease in the threshold electric field strength for the field electron emission. The threshold field strength is one of the main characteristics of efficiency of the field electron emission. This threshold was minimum (less than 1.5 V/μm) for carbon CVD films free of the diamond phase [17–19]. Neither these low threshold values of the macroscopic electric field strength, nor the high densities of the emission current (up to 100 mA/cm^2) and emission

centers (above 10^6 cm^{-2}) or the parameters of current–voltage characteristics of the carbon CVD film cathodes [17–19] can be explained based on the traditional field emission theory [1, 2].

In the previously proposed model, the high efficiency of the field electron emission from carbon CVD films and the low-voltage emission from other carbon materials were explained by the presence of atomic clusters with a modified electron structure in the nanostructured graphite-like carbon phases entering into the composition of such films and some other carbon-based emitters. This nanostructured carbon has the form of nanotubes, platelike graphite nanocrystals, and/or fullerene-like elements [17–19]. In this study we have obtained for the first time direct experimental evidence that the carbon CVD materials contain such clusters characterized by reduced values of the electron work function.

2. EXPERIMENTAL

The carbon film samples were prepared by CVD from a hydrogen–methane ($H_2 : CH_4 \approx 9 : 1$) gas mixture activated by a dc discharge. The experimental film deposition setup was described in detail elsewhere [20]. The carbon CVD films were deposited onto 25×25 mm^2 silicon substrates cut from standard 0.3-mm-thick single crystal silicon wafers. The substrate temperature during deposition was about 1000°C; the deposition process duration was about 90 min.

The field electron emission from carbon film samples was studied in the vacuum diode regime. The anode was a glass plate coated with a transparent conducting indium and tin oxide (ITO) film. The conducting anode film was covered with a phosphor layer. The anode-cathode distance could be varied and adjusted with a micrometric screw drive to within 5 μm. A typical interelectrode spacing d in the course of measurements was 200 μm. The average (macroscopic) electric field strength in the interelectrode gap was calculated as $F = U/d$, where U is the potential difference between anode and cathode. In the plane-parallel electrode configuration, the electric field in the interelectrode gap could be considered as homogeneous, which significantly simplifies analysis of the experimental data as compared to the cases of point [21] or spherical anodes [14, 22].

The current–voltage characteristics were measured using a computer-controlled system, in which the maximum emission current was limited (in order to protect the measuring circuits) on a level of 1 mA. The maximum possible emission current densities were estimated using special experiments in which the main criterion was a long-term stability of the electron emission process. The electron-induced cathodoluminescence from the phosphor layer covering the anode allowed us to evaluate the distribution of emission centers on the cathode surface. The images were photographed or monitored with a video camera.

The Raman scattering spectra were taken and the electron-microscopic images were obtained using standard equipment. The photoelectron spectra were measured on a VG ESCALAB combined spectrometer (Vacuum Generators Scientific, England). The photoelectron spectra were excited either by X-ray radiation with a quantum energy of 1253.6 eV (a characteristic MgK_α emission filtered through an Al foil window) or by rigid UV radiation with a quantum energy of 21.2 eV from a gas-discharge HeI source. The energy scale was calibrated using a freshly cleaved surfaces of gold with $E_b(\mathrm{Au}4f_{7/2})$ = 84.0 eV and copper with $E_b(\mathrm{Cu}2p_{3/2})$ = 932.7 eV [23, 24].

3. RESULTS AND DISCUSSION

The samples of carbon CVD films studied in our experiments were prepared under standard conditions optimized so as to obtain a maximum efficiency of the field electron emission. Stability of the CVD film technology employed provided for a good reproducibility of the results: the parameters measured for several hundred of samples had virtually the same values.

The data of Raman spectroscopy and electron diffraction showed that the CVD film material studied had a graphite-like structure with a sufficiently high degree of crystallinity [17–19]. This conclusion was confirmed by the results of X-ray photoelectron spectroscopy (XPS) measurements. Figure 1 shows a typical review XPS spectrum from which we may conclude, in particular, that the carbon films contain no foreign impurities except for oxygen, the relative bulk concentration of which (i.e., the atomic ratio of oxygen to carbon) was about 1.5 at.% in the analyzed surface layer with a thickness not exceeding 10 nm. Assuming that the presence of oxygen is explained by adsorption during the contact of samples with atmosphere and this oxygen is localized on the film surface, the surface concentration of oxygen may reach up to 5–10%. More detailed conclusions about the structural features of CVD film materials could be derived based on the analysis of the shapes of individual lines in the XPS spectrum.

We performed a comparative analysis of the shape of the C1*s* photoelectron lines in the XPS spectra of carbon CVD films, diamond single crystals, pyrolytic graphite, and fullerene-like carbon, together with the spectra of electron energy losses for the C1*s* line in the same materials. It was established that the C1*s* line shape in the spectrum of CVD carbon is close to that in the spectrum of pyrolytic graphite, whereas significant differences were observed between the former spectrum and the spectra of diamond or fullerene-like graphite. Similar conclusions were derived from an analysis of the spectra of the C *KLL* Auger electron transitions in various carbon materials. An analysis of the shape of these signals is usually employed for the identification of various forms of carbon. Figure 2 shows an example of such spectra for a carbon CVD

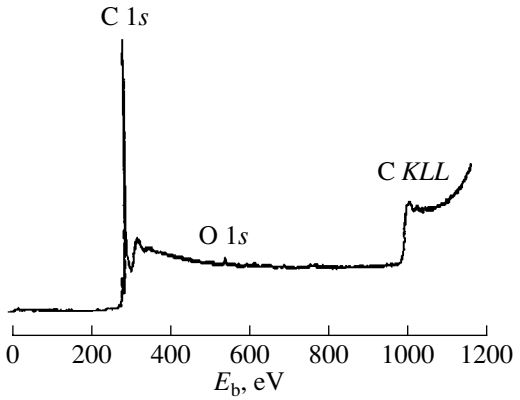

Fig. 1. A review XPS spectrum of a carbon CVD film.

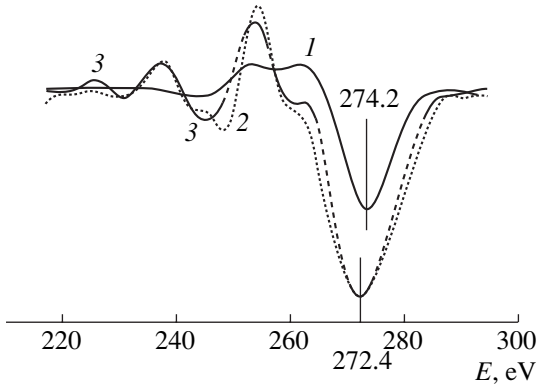

Fig. 2. The C *KLL* Auger electron line in the XPS spectrum of (*1*) diamond, (*2*) pyrolytic graphite, and (*3*) carbon CVD film.

film, pyrolytic carbon, and diamond. As can be seen, the best coincidence of the C *KLL* lines takes pace for the samples of CVD film and graphite.

The type of the carbon material structure is of decisive importance for explaining the emission properties of the films. Figure 3 shows a typical current–voltage (*I–V*) curve measured in the vacuum diode mode for an anode–cathode distance of 200 μm. As can be seen from these data, a threshold macroscopic electric field strength is below 1.5 V/μm, which is several orders of magnitude lower as compared to the values typical of the metal emitters. At the same time, the plot is properly linearized in the Fowler–Nordheim coordinates, which is characteristic of the classical field emitters based on the materials with metal conductivity [1]. Based on the corresponding theoretical relationships, we may use the slope of the linear portion of the *I–V* curve to determine the effective energy barrier for electrons at the emitter–vacuum interface (see, e.g., [3, 14, 17–19, 22]). However, this would require the knowledge of the local electric field strength, which may strongly differ from the macroscopic value as a result of the field concentration on the microscopic points or the other like structures on the cathode surface. Our previous estimates showed that materials possessing an electron work function analogous to that of the conventional graphite are characterized by the field amplification on the order of 5×10^3 to 10^4 [17–19]. However, such a large field amplification contradicts the base assumptions of the Fowler–Nordheim theory formally employed to derive the estimates. Moreover, direct experimental observations of the surface morphology with the aid of electron microscopy give evidence for a much lower role of the geometric factor in the field amplification.

For example, Fig. 4 presents an image of the surface structure of a carbon CVD film examined in a transmission electron microscope (TEM). The film surface region observed in this image contains the formations analogous to the carbon nanotubes and platelike graphite nanocrystals. Using high-resolution TEM measurements, we demonstrated previously [25] that the above nanostructures are composed of the well ordered atomic layers of a graphite-like carbon. A characteristic feature of these structures is the predominant orientation of their atomic layers perpendicularly to the substrate plane. The characteristic diameter of the nanotubes and the grain size is about $D \sim 20$ nm. The film thickness, measured from the substrate to the upper boundary of a dense layer of the platelike graphite crystals is about 4 μm. The length of the nanotubes protrud-

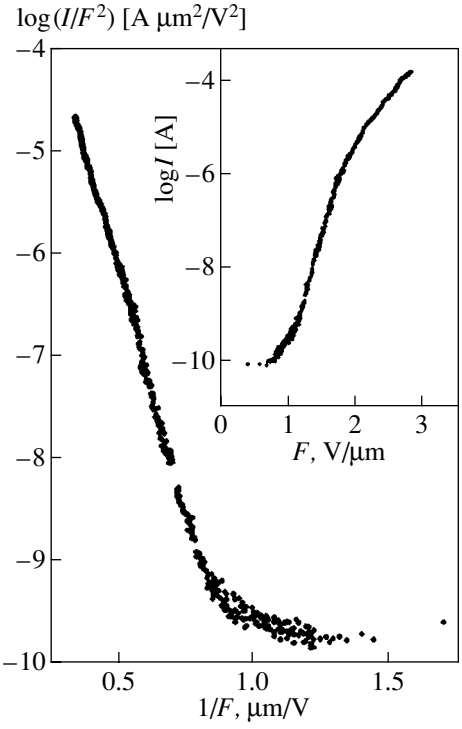

Fig. 3. A typical current–voltage characteristic plotted in the Fowler–Nordheim coordinates for a vacuum diode with cold cathode based on the carbon CVD film. The inset shows the same curve plotted in different coordinates.

Fig. 4. A TEM micrograph of the surface of a graphite-like carbon CVD film.

ing above the film surface usually does not exceed $H \sim 1$ μm.

Based on these characteristic dimensions, the coefficient of the geometric field amplification can be estimated by the order of magnitude as $D/H \sim 50$. Taking into account that the size of the effective emitting area on the nanotube surface may be significantly smaller than the total area, the real field amplification coefficient can be somewhat greater. At the same time, the screening effect of the adjacent tubes must reduce the field amplification. It should be also noted that the shape of the upper part of the nanotubes formed on the carbon CVD film surface may also vary. Besides the so-called open tubes, there can also exist the closed tubes with spherical or cone-shaped tips and various dendrite-like structures (see Fig. 4) [25]. Our observations revealed no correlations between the shape of the nanotubes and nanocrystals in the CVD films and their electron emission properties.

Thus, the field amplification coefficient determined from direct experimental observations turned to be significantly lower as compared to the estimates based on the assumption that the electron work function for a carbon CVD film is close to the value for a usual graphite. In order to determine the possible difference in the electron work function between the CVD films and graphite, we used the method of secondary electron yield spectroscopy.

Figure 5 shows several secondary electron yield spectra measured in the range from 0 to 20 eV. The secondary-electron signal was excited by X-ray radiation and measured by applying a bias potential to a sample in order to separate the electrons emitted from the sample and the analyzer. As is seen in Fig. 5, there are two groups of secondary electrons. The energy shift toward greater kinetic energy for one of these groups exactly corresponds to a bias potential applied to the sample. For the other peak, the shift is about 30% of the corresponding additional energy value. We may suggest that the secondary electrons in the two groups emitted from the sample under the action of the X-ray radiation have different origins. Electrons in the first (high-energy) group are apparently due to excitation of the intrinsic conduction electrons, while the second (low-energy) group is emitted as a result of a two-step process. In the first step, an X-ray quantum is absorbed that leads to the electron–hole pair production. This is followed by the electron emission leading to the formation of a fixed positive charge in a near-surface layer of the sample.

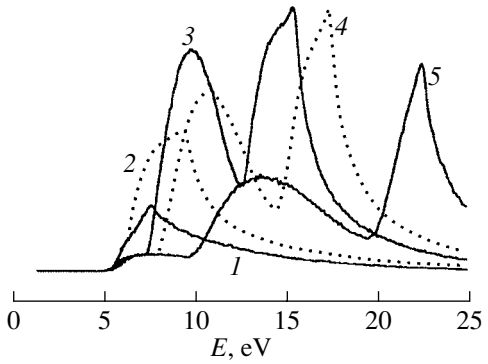

Fig. 5. X-ray-excited secondary electron yield spectra measured for various values of the bias potential applied to a carbon CVD film sample: (*1*) 0; (*2*) –2; (*3*) –8; (*4*) –10; (*5*) –15 V.

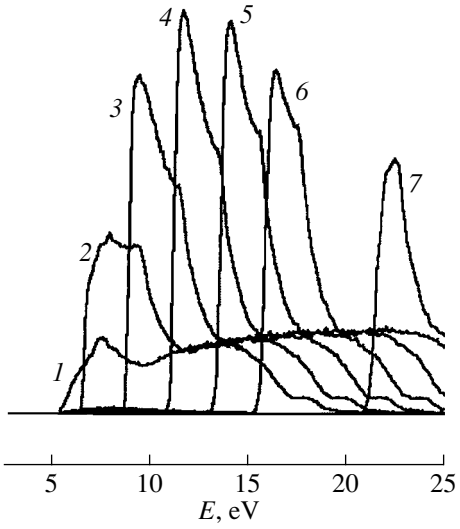

Fig. 6. UV-excited secondary electron yield spectra measured for various values of the bias potential applied to a carbon CVD film sample: (*1*) 0; (*2*) –2; (*3*) –4; (*4*) –6; (*5*) –8; (*6*) –10; (*7*) –15 V.

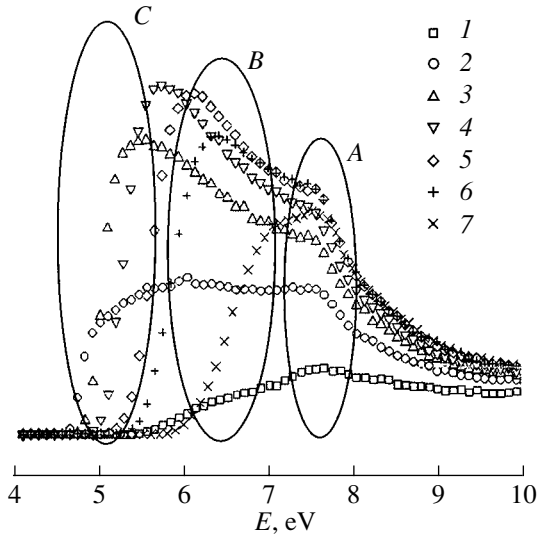

Fig. 7. Low-energy regions of the UV-excited secondary electron yield spectra measured for various values of the bias potential applied to a carbon CVD film sample: (*1*) 0; (*2*) –2; (*3*) –4; (*4*) –6; (*5*) –8; (*6*) –10; (*7*) –15 V. The cutoff energies for all spectra are reduced to a common zero level.

The electric field of the latter positive space charge accounts for the partial decrease in the kinetic energy of electrons emitted from sample into vacuum. In the latter case, the electron emission efficiency is determined by the exciting radiation intensity and by the rate of uncompensated hole recombination. This recombination rate may be rather insignificant if the electron emission and the hole production processes are related to the structures with localized electron densities. The dependence of the emission efficiency on the exciting radiation intensity explains the fact that the UV radiation excites only electrons of the first type, the energy distribution of which shifts precisely in accordance with the applied bias potential (Fig. 6).

A comparative analysis of the secondary electron yield spectra measured using the UV excitation showed that the energy distribution of the secondary electrons emitted from carbon CVD films exhibits a "shoulder" on the low-energy side, which is not observed in the analogous spectrum of crystalline graphite. For better illustration, Fig. 7 shows the low-energy regions of the spectra taken from Fig. 6 and reduced to a common kinetic energy zero. The low-energy shoulder in Fig. 7 corresponds to region *C*, which was observed only in the secondary electron yield spectra of CVD films; the two other characteristic regions (*A* and *B*) were observed in the spectra of all other samples containing a graphite-like carbon fraction. As can be seen in Fig. 7, the low-energy shoulder in the electron energy distribution appears only at a certain threshold value of the bias potential applied to the CVD sample. Lower energies of these secondary electrons correspond to a lower electron work function, while a relatively small fraction of such electrons is indicative either of an insignificant density of the corresponding emitting centers or of their partial screening (which decreases with increasing the bias potential). A nonmonotonic character of the dependence of the low-energy electron fraction on the bias potential may be related to expanded angular distribution that was not monitored in our experiments.

A similar effect was observed in the spectra excited by the X-ray radiation. Figure 8 shows two secondary electron yield spectra comparing a CVD film and a pyrolytic graphite sample measured with the same bias potential of –8 V. By approximating the low-energy region of these spectra with straight lines, we may estimate the corresponding threshold energies P_1 and P_2. These energies proved to be the same for low-energy electrons in the materials of both types. As noted above, these electrons are emitted from the localized states related to structural defects. The cutoff energy of the high-energy electrons emitted from unlocalized states is approximately 1 eV lower in CVD films as compared to the graphite samples. This decrease in the threshold energy is indicative of a reduced electron work function on a certain part of the CVD film surface.

Additional evidence for the presence of surface structures with reduced electron work function is provided by comparison of the secondary emission efficiency, which usually increases with decreasing work function. Figure 9 shows the secondary electron yield spectra for graphite and a CVD film measured with a bias potentials of 0 and –8 eV applied to the samples. In contrast to the graphite sample, the CVD film (nanostructured) material exhibit a significant increase in the secondary electron yield upon application of the additional potential. This result indicates that the structures

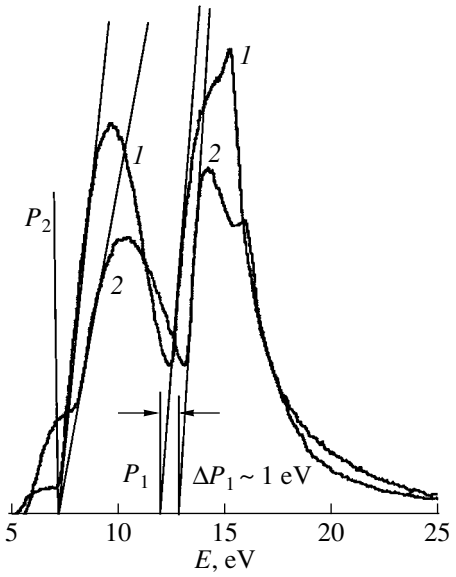

Fig. 8. The X-ray-excited secondary electron yield spectra of (*1*) a carbon CVD film and (*2*) a pyrolytic carbon sample measured for the same bias voltage of –8 V. The figure illustrates the method used for determining the electron cutoff energy.

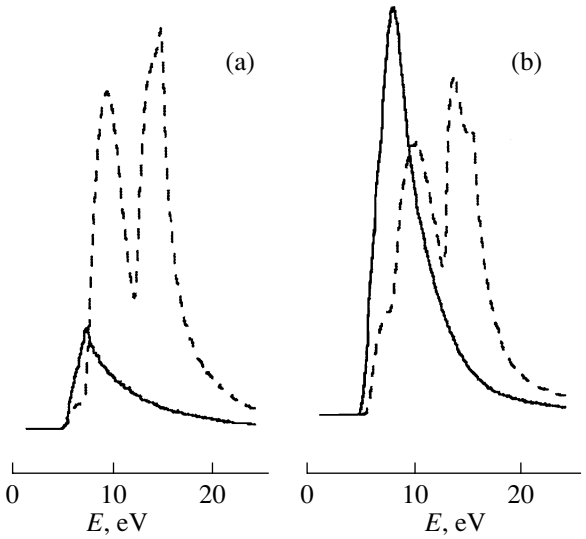

Fig. 9. A comparative of the secondary electron emission yield spectra of (a) a carbon CVD film and (b) a pyrolytic carbon sample measured for a zero bias voltages (solid curve) and for –8 V (dashed curves).

with reduced electron work function are not manifested unless an additional potential is applied, which is related to the existence of a threshold determined by the work function of the electron analyzer in the spectrometer employed. An increase in the secondary electron emission upon application of the additional bias potential may take place in the presence of structures possessing a lower electron work function, and no such change must be observed in the absence of these structures.

Thus, the results of our investigation showed evidence for the presence of elements with reduced electron work function (in comparison to the usual carbon-based materials) in the carbon CVD films. According to the assumptions made previously [18, 19, 25], these structural elements may represent the clusters comprising double atomic chains localized in the bending regions of graphite-like atomic carbon layers. In carbon nanotubes, such chains may be arranged along the side surface. In platelike nanocrystals, these chains may be align in the direction of sharp edges. The presence of such clusters with reduced electron work function provides a self-consistent explanation both for the low-voltage field electron emission and for a number of other phenomena observed in nanostructured carbon materials, including cathodoluminescence [26] and a light emission accompanying the field electron emission [27].

According to the above assumptions, an emission center on the surface of a nanostructured carbon material can represent a single bent atomic layer or several such layers (Fig. 10). The graphite bonds between most part of the carbon atoms in such structural elements provide for the high electric conductivity. The presence of clusters with modified electron properties on the graphite surface accounts for a change in the energy barriers at the emitter–vacuum interface, which is illustrated by the energy band diagram in Fig. 10. The density of electron states for atoms in the clusters must differ significantly from that characteristic of a semimetallic graphite with a zero bandgap.

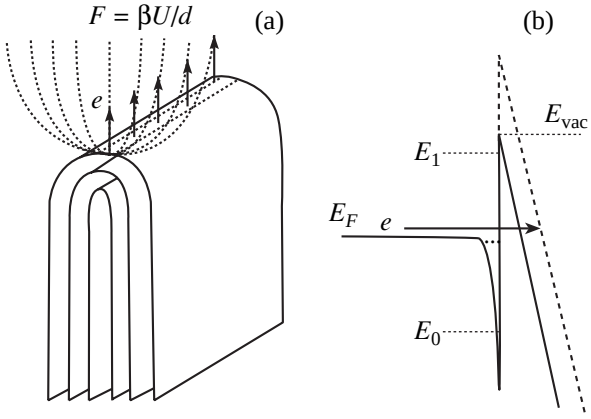

Fig. 10. Schematic diagrams illustrating (a) the proposed model of the emission center in a nanostructured carbon material and (b) the corresponding energy band diagram in the presence of an applied electric field $F = \beta U/d$ ($\beta = D/H$); E_0 and E_1 are the energy levels bounding the bandgap in the electron density of states of the carbon clusters; E_{vac} is the vacuum level; E_F is the Fermi level. Dashed lines show the energy band diagram of the usual graphite.

The most important difference is the appearance of a bandgap in the density of states, which is bounded from above and below by the energy levels E_1 and E_0, respectively. As indicated previously [17–19] it is natural to suggest that the width of this bandgap is about 4 eV. When an external electric field is applied to this heterogeneous structure, electrons are injected from the graphite-like region into a near-surface layer possessing characteristics equivalent to those of a wide-bandgap material, the surface of which may be expected to exhibit a low or negative electron affinity (analogous to those observed in diamond). A thickness of the transition region between the graphite-like and diamond-like parts of this emitting center does not exceed several lattice constants, which accounts for the effective electron transport in this structure.

The proposed model of the emitting center and the mechanism of the field electron emission are not specific of the CVD films of the particular type studied. Analogous mechanisms can explain the other experimental observations of the low-voltage electron emission from various carbon-based materials composed of or containing nanostructured graphite-like elements in the form of nanotubes and/or nanocrystals. Besides the polycrystalline diamond films, diamond–graphite compounds [5, 7, 9–16], and carbon nanotubes as such [22, 27], materials of this type may include the amorphous carbon films composed of fullerene-like elements [28], carbon fibers [29, 30], and powdered [31] or disordered graphite [32], etc.

4. CONCLUSION

The experimental data presented above confirm our previous suggestions concerning the presence of clusters with diamond-like interatomic bonds in nanostructured graphite-like carbon materials. A local decrease in the electron work function in the region of such clusters favors a decrease in the threshold value of the electric field strength for which the field electron emission takes pace. The proposed model of the emission center and the mechanism of the field electron emission explain a number of effects observed in various carbon materials, which are indicative of important distinctions from the field emission phenomena observed for metal cathodes.

REFERENCES

1. R. G. Fowler and L. Nordheim, Proc. R. Soc. London, Ser. A **119** (781), 173 (1928).
2. *Non-Incandesced Cathodes*, Ed. by M. I. Elinson (Sovetskoe Radio, Moscow, 1974).
3. C. A. Spindt, I. Brodie, L. Humphrey, and E. R. Westerberg, J. Appl. Phys. **47**, 5248 (1976).
4. J. Robertson, Thin Solid Films **296**, 61 (1997).
5. R. J. Nemanich, P. K. Baumann, M. C. Banjamin, *et al.*, Diamond Relat. Mater. **5**, 790 (1996).
6. F. J. Himsel, J. A. Knapp, J. A. van Vechten, and D. E. Eastman, Phys. Rev. B **20**, 624 (1979).
7. M. W. Geis, N. N. Efremov, J. D. Woodhouse, *et al.*, IEEE Electron Device Lett. **12**, 456 (1991).
8. S. A. Kajihara, A. Antonelli, J. Bernholc, and R. Car, Phys. Rev. Lett. **66**, 2010 (1991).
9. M. W. Geis, N. N. Efremov, K. E. Krohn, *et al.*, Nature **393**, 431 (1998).
10. W. Zhu, G. P. Kochanski, S. Jin, *et al.*, Appl. Phys. Lett. **67**, 1157 (1995).
11. J. W. Steeds, A. Gilmore, K. M. Bussman, *et al.*, Diamond Relat. Mater. **8**, 996 (1999).
12. A. T. Sowers, B. L. Ward, S. L. English, *et al.*, Diamond Relat. Mater. **9**, 1569 (2000).
13. W. Zhu, G. P. Kochanski, S. Jin, and L. Seibles, J. Appl. Phys. **78**, 2707 (1995).
14. O. Gröning, O. M. Küttel, P. Gröning, and L. Schlapbach, J. Vac. Sci. Technol. B **17**, 1970 (1999).
15. J. Ristein, Diamond Relat. Mater. **9**, 1129 (2000).
16. V. Ralchenko, A. Karabutov, I. Vlasov, *et al.*, Diamond Relat. Mater. **8**, 1496 (1999).
17. A. N. Obraztsov, V. P. Volkov, and I. Yu. Pavlovskiĭ, Pis'ma Zh. Éksp. Teor. Fiz. **68**, 56 (1998) [JETP Lett. **68**, 59 (1998)].
18. A. N. Obraztsov, I. Yu. Pavlovsky, and A. P. Volkov, J. Vac. Sci. Technol. B **17**, 674 (1999).
19. A. N. Obraztsov, A. P. Volkov, and I. Yu. Pavlovsky, Diamond Relat. Mater. **9**, 1190 (2000).
20. I. Yu. Pavlovskiĭ and A. N. Obraztsov, Prib. Tekh. Éksp., No. 1, 152 (1998).
21. V. D. Frolov, A. V. Karabutov, V. I. Konov, *et al.*, J. Phys. D **32**, 815 (1999).
22. O. Gröning, O. M. Küttel, Ch. Emmeneger, *et al.*, J. Vac. Sci. Technol. B **18**, 665 (2000).
23. J. E. Moulder, W. F. Stickle, P. E. Sobol, and K. D. Bomben, *Handbook of X-ray Photoelectron Spectroscopy* (Perkin-Elmer, Eden Prairie, 1992).
24. *Practical Surface Analysis by Auger and X-ray Photoelectron Spectroscopy*, Ed. by D. Briggs and M. Seah (Wiley, New York, 1983; Mir, Moscow, 1987).
25. A. N. Obraztsov, A. P. Volkov, I. Yu. Pavlovskiĭ, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 381 (1999) [JETP Lett. **69**, 411 (1999)].
26. A. P. Volkov, A. N. Obraztsov, I. Yu. Pavlovskiĭ, *et al.*, Poverkhnost, Nos. 5–6, 161 (1999).
27. J.-M. Bonard, T. Stockli, F. Maier, *et al.*, Phys. Rev. Lett. **81**, 1441 (1998).
28. H.-J. Ssheibe, H. Banzhof, A. Luft, *et al.*, Abstract No. 8B.4 of the International Conference DIAMOND'98, Greece, 1998.
29. C. Lea, J. Phys. D **6**, 1105 (1973).
30. E. P. Sheshin, Ultramicroscopy **79**, 101 (1999).
31. A. Y. Tcherepanov, A. G. Chakhovskoi, and V. B. Sharov, J. Vac. Sci. Technol. B **13**, 482 (1995).
32. A. L. Suvorov, E. P. Sheshin, V. V. Protasenko, *et al.*, Vide, Couches Minces, No. 271, 326 (1994).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 93, No. 4, 2001, pp. 853–859.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 979–985.

SOLIDS
Electronic Properties

Plasma Oscillations in Nanotubes and the Aharonov–Bohm Effect for Plasmons

A. I. Vedernikov, A. O. Govorov, and A. V. Chaplik*

Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. akademika Lavrent'eva 13, Novosibirsk, 630090 Russia

**e-mail: chaplik@isp.nsc.ru*

Received April 12, 2001

Abstract—We theoretically analyze the collective oscillations of 2D electrons in nanotubes. In the presence of a magnetic field parallel to the tube axis, the plasmon frequencies undergo Aharonov–Bohm oscillations. The effect can manifest itself in infrared absorption and in Raman scattering. We calculate the cross sections for inelastic light scattering by plasmons. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Nanotubes (and quantum rings) occupy a prominent place among the objects studied by modern physics of low-dimensional systems because of their topological peculiarities. Since the region of electron motion is not simply connected, peculiar effects result (in the presence of a magnetic field), in which the phase of the wave function proves to be observable. All these effects are the derivatives of the well-known Aharonov–Bohm effect.

In a magnetic field parallel to the nanotube axis, the single-particle spectrum depends on magnetic flux as

$$E_m(q) = \frac{\hbar^2 q^2}{2\mu} + B(m+\phi)^2, \quad B = \frac{\hbar^2}{2\mu a^2}, \qquad (1)$$

where μ is the effective mass, a is the cylinder radius, $\hbar q$ is the momentum along the axis, ϕ is the number of magnetic-flux quanta inside the tube, and $m = 0, \pm 1, \pm 2, \ldots$ is the azimuthal quantum number. This dependence of E_m on ϕ results in oscillations in macroscopic properties of the nanotube, for example, in conductivity [1] or magnetic moment [2]. In both cases, macroscopic manifestations of the properties of charge-carrying elementary excitations that obey Fermi statistics or, more simply, the electron behavior are considered.

Recently, a number of authors [3] have shown that Aharonov–Bohm oscillations also take place for a neutral object—an exciton in a quantum ring. The possibility of an electron and a hole tunneling toward each other along the ring leads to oscillatory dependencies of the exciton binding energy and formation probability on magnetic flux.

In all the above examples, the oscillation period is universal and equal to the flux quantum $\phi_0 = hc/e$. However, as was shown by Chaplik [4], this universality breaks down for a charged exciton (trion). Its binding energy oscillates with flux with a period that depends on the ratio of the effective electron and hole masses.

It seems of interest to explore the possibility of oscillation effects in the collective excitations of an electron system in nanotubes. Here, we show that Bose neutral elementary excitations (plasmons) are also characterized by oscillatory dependencies of their parameters on magnetic flux, i.e., they exhibit the Aharonov–Bohm effect. This effect manifests itself in nanotube optical properties and can thus be observed in experiments that require no electrical contact with the objects being studied.

We consider two types of nanotubes: semiconductor hollow cylinders (for example, self-assembled quantum rolls [5]) and carbon nanotubes. For the former, we use a standard parabolic dispersion law of two-dimensional electrons that leads to formula (1) in a magnetic field. For the latter, we take a conical dispersion law of two-dimensional graphite [6] as the initial one:

$$E^{\pm} = \pm \hbar V_0 \mathbf{q},$$

where $\mathbf{q}$ is the two-dimensional (2D) vector in the plane of the graphite sheet, and V_0 is a parameter of the order of the electron velocity in the atom.

Below, we derive the dispersion law for plasma oscillations, the dependence of the plasmon frequency on magnetic flux in a magnetic field parallel to the nanotube axis, the infrared absorption spectrum, and the cross section for inelastic light scattering by plasmons in nanotubes.

2. PLASMA OSCILLATIONS FOR $\phi = 0$; THE COLD-PLASMA APPROXIMATION

In the simplest theory of plasma oscillations, the spatial dispersion is ignored, which is equivalent to ignoring the particle velocity distribution (cold plasma). The corresponding criterion for a degenerate

plasma is $\omega \gg kV_F$, where ω and k are the frequency and wave vector of the plasma wave, and V_F is the Fermi velocity. In this case, the system of equations for plasma oscillations (without retardation effects) is

$$\begin{aligned} \Delta\varphi &= -4\pi e\delta(\rho - a)\tilde{N}_s, \\ e\dot{\tilde{N}}_s + \mathrm{div}\mathbf{j}_s &= 0, \quad \mathbf{j}_s = -\sigma\nabla\varphi_{\rho = a}. \end{aligned} \tag{2}$$

Here, $\tilde{N}_s$ and $\mathbf{j}_s$ are an addition to the particle surface density and the surface current, respectively; σ is the two-dimensional conductivity; and φ is the electric potential. For parabolic dispersion and in the collisionless approximation,

$$\sigma = \frac{iN_s e^2}{\mu\omega},$$

where N_s is the equilibrium electron surface density. Solving system (2) in cylindrical coordinates yields the dispersion of a plasmon with momentum k along the nanotube axis and with azimuthal moment m:

$$\omega_p^2 = \frac{4\pi e^2 N_s}{\mu} a\left(k^2 + \frac{m^2}{a^2}\right) K_m(ka) I_m(ka), \tag{3}$$

where I_m and K_m are, respectively, the Bessel functions of the first and third kinds of an imaginary argument. The dispersion law (3) gives the correct asymptotics.

(i) In the long-wave limit, $ka \ll 1$, for an axially symmetric plasmon ($m = 0$):

$$\omega_p^2 = \frac{2e^2 N_L k^2}{\mu}\ln\frac{2}{ka\gamma}, \quad \gamma = e^{-C}, \tag{4}$$

where C is the Euler constant, and $N_L = 2\pi a N_s$ is the linear electron density;

(ii) In the shortwave limit, $ka \gg 1$, $m \gg 1$:

$$\omega_p^2 = \frac{2\pi e^2 N_s}{\mu}\sqrt{k^2 + \frac{m^2}{a^2}}, \tag{5}$$

which corresponds to a 2D plasmon with the momentum components (k, m/a).

Equation (4) represents a standard dispersion law for a one-dimensional plasmon (for example, in a quantum wire), which usually includes the cutoff size under the logarithm known only in order of magnitude. As we see, the result for a nanotube is completely definite, including the numerical coefficient under the logarithm.

The logarithmic singularity of $\omega(k)$ when $m = 0$ and $k \longrightarrow 0$ formally corresponds to an infinite group velocity, which is, of course, unfeasible. At low k, the retardation effects (transverse fields) must be taken into account. This requires solving the Maxwell equations for the scalar and vector potentials instead of system (2):

$$\begin{aligned} \Delta\varphi - \frac{1}{c^2}\frac{\partial^2\varphi}{\partial t^2} &= -4\pi e\tilde{N}_s\delta(\rho - a), \\ \Delta\mathbf{A} - \frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2} &= -\frac{4\pi}{c}\mathbf{j}_s\delta(\rho - a), \\ \mathbf{j}_s &= -\sigma\left(\nabla\varphi + \frac{1}{c}\frac{\partial\mathbf{A}}{\partial t}\right)_{\rho = a}. \end{aligned} \tag{6}$$

As a result, the dispersion law is given by Eq. (3), in which the following substitution must be made:

$$k \longrightarrow R \equiv \sqrt{k^2 - \frac{\omega^2}{c^2}}.$$

For $m = 0$ at $ka \ll 1$, we have

$$\omega_p^2 = \frac{2e^2 N_L k^2 \Lambda}{\mu(1 + 2e^2 N_L \Lambda/\mu c^2)}, \quad \Lambda = \ln\frac{2}{Ra\gamma}. \tag{7}$$

Thus, we have $\omega \approx ck$ when $k \longrightarrow 0$, but the domain of existence of this asymptotics is exponentially small, of the order of

$$\exp\left(-\frac{\mu c^2}{2e^2 N_L}\right).$$

Let us now consider the dispersion with a conical point. Depending on the method of rolling up the graphite sheet, the nanotube can have either semiconductor or metal band structure. In the latter case, the gap in the single-particle spectrum vanishes at $q = 0$, while the density of states remains finite at this point; the azimuthal quantum number of the electron is zero, and $E_0^{\pm} = \pm\hbar V_0|q|$, where the "+" and "–" signs refer to the conduction and valence bands, respectively.

Clearly, for a nonzero gap in the electron spectrum, the pattern of plasma oscillations does not differ qualitatively from that for the parabolic dispersion considered above. Therefore, we consider in more detail only the metal band structure, for which nontrivial singularities appear in the plasmon parameters.

We assume that only the zeroth (in azimuthal quantum number) one-dimensional subband is occupied. For a degenerate system, this implies a constraint either on the dopant density,

$$n_L < \frac{2}{\pi a},$$

or on the donor energy level (which is positive for conical dispersion),

$$E_d < \frac{\hbar V_0}{a}.$$

For an intrinsic conductivity, the temperature must be low enough, $T \ll \hbar V_0/a$. An expression for the nanotube conductivity can be easily derived from the collisionless kinetic equation if the dispersion law is linear:

$$\sigma = \frac{ie^2 V_0}{\pi^2 \hbar a \omega}. \quad (8)$$

All relations (2) hold, and we find the plasmon frequency for degenerate electrons to be

$$\omega_p^2 = \frac{4e^2 V_0 k^2}{\pi\hbar} K_0(ka) I_0(ka), \quad \omega_p \ll kV_0. \quad (9)$$

It can be shown that the square of the plasma frequency for an arbitrary ω_p/kV_0 is $k^2 V_0^2 + \omega_p^2$, with ω_p^2 defined in (9).

This is the case for doped carbon nanotubes at a zero temperature. Without doping and at a finite temperature, there is an intrinsic two-band conductivity. The quantity ω_p^2 is proportional to $f(0)$, the value of the Fermi function at $E = 0$. $f(0) = 1$ in a degenerate system at a zero temperature and $f(0) = 1/2$ for a nanotube with an intrinsic conductivity, because the chemical potential is zero at any temperature (the conical dispersion law!). The factor 1/2 is compensated for, because electrons and holes contribute equally to the conductivity, and we again derive a plasmon dispersion law in the form (9). Its characteristic (and, at first glance, paradoxical) feature is that the plasma frequency is independent of the carrier density. The same is also true for the conductivity (8). The reason can be easily understood if we note that both ω_p^2 and σ are proportional to N_L/μ for parabolic electron dispersion. Conical dispersion can be formally obtained if the effective mass μ itself is assumed to depend linearly on q. In that case, its value at the Fermi level (degenerate gas) or the mean temperature value (nondegenerate gas) enter into the formulas for ω_p^2 and σ. The parameter N_L matches the Fermi wave vector, to within a numerical factor, and is proportional to $T/\hbar V_0$ in a nondegenerate system. Thus, in both cases, the dependence on N_L is eliminated from the formulas. This is because the density of states in a one-dimensional subband is constant for a zero azimuthal number.

It is well known that carbon nanotubes can form coaxial structures with different numbers of nested cylinders. The probability of electron transitions between them is negligible, but the coupling through electric fields of plasma oscillations causes the number of branches of the plasmon spectrum to increase (an analog of a planar multilayer structure). For example, in the case of two coaxial nanotubes, two branches corresponding to in-phase and out-of-phase oscillations (optical and acoustic plasmons) emerge:

$$\omega_\pm^2 = \frac{\omega_{pa}^2 + \omega_{pb}^2}{2} \pm \sqrt{\frac{(\omega_{pa}^2 - \omega_{pb}^2)^2}{4} + \omega_{pa}^2 \omega_{pb}^2 \Lambda_m}. \quad (10)$$

Here,

$$\Lambda_m(k; a, b) = \frac{I_m(ka)K_m(kb)}{I_m(kb)K_m(ka)} < 1,$$

a and b are the radii of the two nanotubes with $b > a$, and ω_{pa} and ω_{pb} are their individual plasma frequencies. The "+" sign before the square root in (10) corresponds to an optical plasmon, and the long-wave asymptotics $\omega_+ \propto k\sqrt{|\ln k|}$ at $m = 0$. The second branch for the ka, $kb \ll 1$, $m = 0$, obeys a linear dispersion law, $\omega_- \propto k\sqrt{\ln(b/a)}$. At $m \neq 0$, both frequencies ω_+ and ω_- tend to the constant values that correspond to intersubband transitions with a depolarization shift (see below) when $k \longrightarrow 0$.

3. EFFECTS OF A MAGNETIC FIELD

A magnetic filed can affect the plasmon dispersion law only via the constitutive equations (current–field relation). If, as was done in Section 2, the conductivity is assumed to be classical, then a longitudinal magnetic field for parabolic dispersion cannot change the orbital motion of 2D electrons on the cylinder surface, and, therefore, σ does not depend on the field. The influence of a magnetic field (or, to be more precise, magnetic flux ϕ) shows up only when quantum effects are taken into account. The dependence of energy on ϕ given by Eq. (1) transforms to a dependence of the Fermi level and polarization operator on magnetic flux. With an allowance for the spatial dispersion of conductivity, an oscillatory dependence of the plasma frequency on ϕ (the Aharonov–Bohm effect for plasmons) arises.

We begin with a symmetric plasmon with $m = 0$. Calculating the polarization operator

$$\Pi(k, m = 0) = \sum_{qm'} \frac{f'(E_{m'}(q))\hbar kq/\mu}{\omega + i\delta - \hbar kq/\mu} \quad (11)$$

and solving the Poisson equation yields the dispersion equation

$$1 = \frac{2e^2 k^2}{\pi\hbar} K_0(ka) I_0(ka) \sum_{m'} \frac{V_F(m')}{\omega^2 - k^2 V_F^2(m')}, \quad (12)$$

where

$$V_F(m') = \sqrt{\frac{2}{\mu}[E_F - B(m' + \phi)^2]},$$

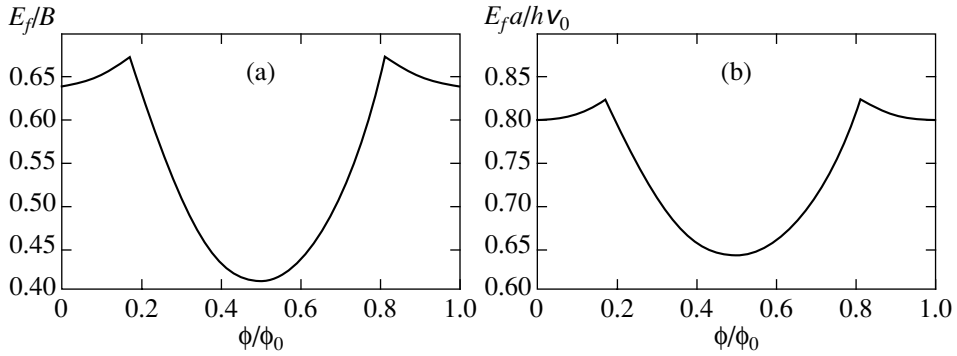

Fig. 1. Fermi energy of degenerate electrons on the nanotube surface versus magnetic flux: (a) a parabolic dispersion law and (b) a conical dispersion law.

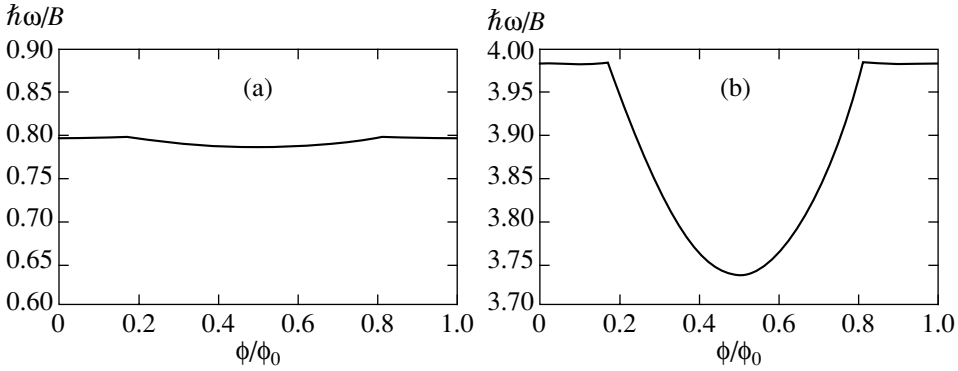

Fig. 2. Plasmon frequency at $m = 0$ versus magnetic flux (parabolic dispersion): $ka = 0.1$ (a) and 1.0 (b).

the Fermi energy must be derived from the equation

$$\hbar N_L = \frac{2\sqrt{2\mu}}{\pi}\sum_{m'}\sqrt{E_F - B(m' + \phi)^2}. \qquad (13)$$

The summation over m' in (12) and (13) is constrained by the requirement that the radicands be positive. An analytic (and relatively simple) answer can be obtained in the approximation of weak spatial dispersion, $\omega \gg kV_F(m')$ for all admissible m'. Expanding in (12) in kV_F/ω and applying the Poisson summation formula yields

$$\omega^2 \approx \frac{2e^2k^2N_L}{\mu}K_0(ka)I_0(ka) + \frac{3(k\bar{V}_F)^2}{2\pi^2N_L a} \times \sum_{l=-\infty}^{\infty} J_2\left(2\pi l\sqrt{\frac{\bar{E}_F}{B}}\right)\frac{\cos(2\pi l\phi)}{l^2}. \qquad (14)$$

Here,

$$\bar{E}_F = \frac{\mu\bar{V}_F^2}{2} = \frac{2\pi\hbar^2N_s}{\mu}$$

is the oscillation-averaged Fermi level. The second term in (14) has a relative smallness of the order of $a^*/a|\ln(ka)|$, where $a^* = \hbar^2/\mu e^2$ is the effective Bohr radius (this is the parameter $k^2V_F^2/\omega^2$ for ω of the order of the frequency of one-dimensional plasma oscillations, which is given by the first term in Eq. (14)). Thus, when spatial dispersion is taken into account, the plasmon frequency oscillates with magnetic flux with the period $\Delta\phi = 1$.

For an arbitrary value of kV_F/ω, we calculated $\omega(\phi)$ numerically. Figure 1 shows a plot of Fermi energy against magnetic flux within one period. The electron density was chosen in such a way that no more than two subbands were occupied at any ϕ (for $\phi > 0$, these are the pairs $m' = 0$ and $m' = -1$; $m' = -1$ and $m' = -2$; and so on). Under the same conditions, Fig. 2 shows $\omega(\phi)$ for two values of the wave vector. As must be the case, the magnetic dispersion of the plasmon frequency is enhanced with increasing ka.

The results are qualitatively different for conical quasi-particle dispersion. In this case, including a magnetic field opens a gap in the dispersion law for a nanotube with a "metal-type" spectrum:

$$E_0^{\pm} = \hbar V_0\sqrt{q^2 + \frac{\phi^2}{a^2}}. \qquad (15)$$

As a result, the diagonal velocity element along the axis and the nanotube classical conductivity depend on magnetic flux even in the lowest approximation in kV_0/ω:

$$\sigma = \frac{ie^2 N_L V_0}{\pi\hbar\omega}\frac{1}{\sqrt{(\pi a N_L)^2 + \phi^2}}. \quad (16)$$

This leads to a more complex dependence of the plasma frequency on ϕ, as illustrated by Fig. 3 for three values of ka. The linear density N_L again corresponds to the occupation of no more than two subbands.

Azimuthally nonuniform oscillations ($m \neq 0$) are similar to intersubband 2D plasmons in quantum films. These are in fact transitions between subbands, $m' \longrightarrow m' + m$, with Coulomb effects (depolarization shift). Since the general case with $k \neq 0$ and $m \neq 0$ is described by cumbersome formulas, we restrict our analysis to a purely transverse plasmon ($k = 0$). The corresponding equation for Π in the case of parabolic dispersion is

$$\Pi(k = 0, m) = \frac{\sqrt{2\mu}m^2}{\pi^2 a\hbar B} \times \sum_{m'} \frac{\sqrt{E_F - B(m' + \phi)^2}}{m^4 - (2mm' + 2m\phi - \omega/B)^2}, \quad (17)$$

where E_F can be determined from Eq. (13). Π is clearly periodic in ϕ with the same period $\Delta\phi = 1$. Figure 4 shows plots of the intersubband-plasmon frequencies against magnetic flux for $m = \pm 1$ in the range $0 < \phi < 1/2$, which corresponds to one half-period [E_F is an even function of ϕ, $\Pi(\phi, \omega) = \Pi(-\phi, -\omega)$].

The density N_L was chosen in such a way that only the $m' = 0$ and $m' = -1$ subbands were occupied as ϕ varied over the above range; this requires that the condition $\pi N_L a < 2$ be satisfied. The break in the plot corresponds to the onset of the $m' = -1$ subband occupation.

For conical dispersion, the case with $m \neq 0$ is qualitatively similar to that considered above.

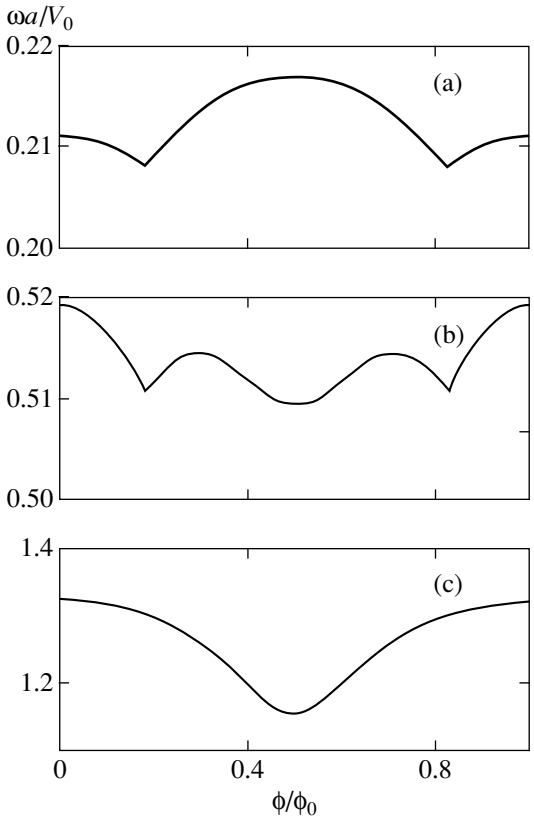

Fig. 3. Plasmon frequency at $m = 0$ versus magnetic flux (conical dispersion); $ka = 0.1$ (a), 0.3 (b), and 1.0 (c).

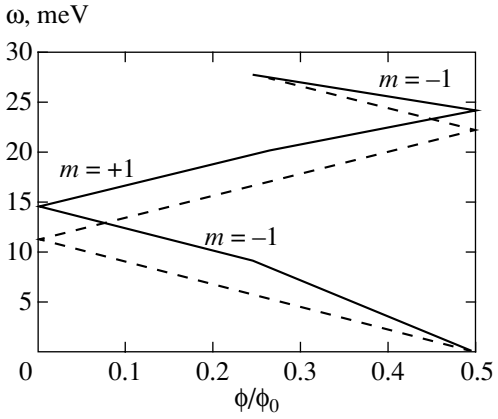

Fig. 4. Frequencies of single-particle transitions (dashed lines) and intersubband plasmons with a depolarization shift (solid lines) versus magnetic flux; $a = 70$ Å and $N_L = 3 \times 10^5$ cm^{-1}.

4. INTERACTION OF PLASMONS WITH ELECTROMAGNETIC RADIATION

4.1. Infrared Absorption

An electromagnetic wave linearly polarized along the nanotube axis produces an axially symmetric perturbation ($m = 0$). This wave can be absorbed by $m = 0$ plasmons if its electric field is modulated in the axis direction with a period L. This is usually achieved with a one-dimensional grating structure. The absorption-line frequency is then equal to the frequency of a plasmon with $k = 2\pi/L$. Given the electron scattering, the absorption at the line center per unit surface area is [7] $E_0^2\sigma_0/2$, where E_0 is the amplitude of the wave electric field, and $\sigma_0 = e^2 N_s\tau/\mu$ is the nanotube static surface conductivity (N_s is the surface density, and τ is the momentum relaxation time).

No modulation is required when the wave is polarized perpendicular to the tube axis: a uniform field excites the $\Delta m' = \pm 1$ transitions; i.e., intersubband plasmons with $m = \pm 1$ are produced in the system. With an

allowance for the depolarization effects, absorption frequency is plotted against ϕ in Fig. 4.

In conclusion, recall that the above discussion pertains to intraband transitions. For carbon nanotubes, these transitions are as follows:

$$E^+_{m'}(q) \longrightarrow E^+_{m'+m}(q+k).$$

The absorption that is attributable to $E^+ \longrightarrow E^-$ transitions and that is unrelated to plasma effects was considered in [8]. The Aharonov–Bohm effect that results from a periodic dependence of the ground-state energy on magnetic flux also shows up in this case.

4.2. Inelastic Light Scattering

The plasmon contribution to inelastic light scattering is commonly observed in the geometry of parallel polarizations of the incident ($\mathbf{e}_1$) and scattered ($\mathbf{e}_2$) photons; the amplitude of the process is proportional to $\mathbf{e}_1 \cdot \mathbf{e}_2$. The corresponding cross section is determined by the density-density correlator. The latter, in turn, is related to the generalized susceptibility α, which gives a response of the system to a scalar perturbation of the form $\exp(i\boldsymbol{\kappa} \cdot \mathbf{r})$ with allowance for the self-consistent field.

For a fixed transfer of the photon momentum with components $\mathbf{k}_\perp$ (in a plane perpendicular to the nanotube axis) and k (along the axis), the scattering cross section is the sum of partial contributions $\sigma_m(k)$ whose relative weights are determined by the coefficients of the expansion of a plane wave $\exp(i\mathbf{k}_\perp \cdot \boldsymbol{\rho})$ in terms of cylindrical harmonics:

$$\frac{d^2\sigma_m}{d\Omega d\omega} = 2al\frac{\omega_1}{\omega_2}\left(\frac{e^2}{m_0c^2}\right)^2 (n(\omega)+1) \times |J_m(k_\perp a)|^2 \mathrm{Im}\alpha_m(k,\omega)(\mathbf{e}_1\cdot\mathbf{e}_2)^2 F(\omega_1), \quad \omega = \omega_1 - \omega_2, \tag{18}$$

where the subscripts 1 and 2 denote, respectively, the incident and scattered photons; l is the nanotube length; m_0 is the mass of a free electron; $n(\omega)$ are the Bose occupation numbers, and J_m is the Bessel function. We also included the amplification factor $F(\omega_1)$ in the formula, because Raman scattering is commonly observed at resonance, when the frequency ω of the exciting light is close to a particular interband transition. In A_3B_5 semiconductors for resonance with a spin-orbit split-off band (see [9]),

$$F = |p_{cv}|^4/9m_0^2\Delta^2,$$

where p_{cv} is the interband momentum matrix element, and Δ is the resonance detuning. The partial susceptibility $\alpha_m(k,\omega)$, which gives a response of the density to a scalar perturbation $f_m e^{im\varphi}$ with moment m, can be calculated by adding $-\nabla f/e$ to the electric field in the formula for current $\mathbf{j}_s$ and by solving system (6). As a result, we obtain

$$\alpha_m(k,\omega) = \frac{i\sigma\left(k^2+\frac{m^2}{a^2}\right)}{e^2\left[\omega + 4\pi i\sigma\left(R^2+\frac{m^2}{a^2}\right)aK_m(Ra)I_m(Ra)\right]}. \tag{19}$$

This expression for α corresponds to the cold-plasma approximation,

$$\omega \gg V_F\sqrt{k^2+\frac{m^2}{a^2}}.$$

Apart from the plasmon pole, $\alpha_m(k,\omega)$ has a bifurcation point at $R = 0$, i.e., at $\omega = ck$. As in the two-dimensional case (see [10]), the high-frequency wing in the Raman scattering spectrum at $\omega > ck$ corresponds to this bifurcation. The imaginary part of α, which gives the wing intensity distribution, is

$$\mathrm{Im}\alpha_m(k,\omega) = 2\pi^2(N_s k/\mu)^2 a(\omega^2/c^2 - k^2)J_m^2 \times\left\{\left[\omega^2 - u^2\left(\frac{\omega^2}{c^2}-k^2\right)J_m N_m\right]^2 + \left[u^2\left(\frac{\omega^2}{c^2}-k^2\right)J_m^2\right]^2\right\}^{-1}, \quad \omega \geq ck, \tag{20}$$

where

$$u^2 \equiv \frac{2\pi^2 e^2 a N_s}{\mu}$$

and the argument of all Bessel functions is $a\sqrt{\omega^2/c^2 - k^2}$.

Intersubband transitions ($m \neq 0$) are excited during inelastic light scattering with $k_\perp \neq 0$. At $k_\perp a \ll 1$, the cross sections for such processes rapidly decrease with increasing m: $\sigma_m \propto (k_\perp a)^{2m}$. In the presence of a magnetic field, the polarization operator should be substituted for the classical conductivity σ in Eq. (19):

$$\left(k^2+\frac{m^2}{a^2}\right)\sigma \longrightarrow ie^2\omega\Pi_m(k). \tag{21}$$

In that case, the peaks in the partial cross sections as functions of ω are determined by resonances on intersubband plasmons and shift as the magnetic flux changes (see Fig. 4). The amplitudes of these peaks, i.e., the partial cross sections themselves, are also periodic functions of ϕ.

Thus, we have shown that the basic parameters of plasma waves in nanotubes oscillate with magnetic flux with a period ϕ_0. For a longitudinal, axially symmetric

plasmon, this dependence emerges only beginning with terms of the order of $(\omega/kV_F)^2$ in the dispersion law.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project no. 99-02-17127) and the Program "Physics of Solid-State Nanostructures." Two of us (A.I.V. and A.O.G.) thank the Volkswagen-Stiftung Foundation (Germany) for support.

REFERENCES

1. B. L. Al'tshuler, A. G. Aronov, and B. Z. Spivak, Pis'ma Zh. Éksp. Teor. Fiz. **33**, 101 (1981) [JETP Lett. **33**, 94 (1981)]; D. Yu. Sharvin and Yu. V. Sharvin, Pis'ma Zh. Éksp. Teor. Fiz. **34**, 285 (1981) [JETP Lett. **34**, 272 (1981)].
2. I. O. Kulik, Pis'ma Zh. Éksp. Teor. Fiz. **11**, 407 (1970) [JETP Lett. **11**, 275 (1970)].
3. A. V. Chaplik, Pis'ma Zh. Éksp. Teor. Fiz. **62**, 885 (1995) [JETP Lett. **62**, 900 (1995)]; R. A. Römer and M. E. Raikh, Phys. Rev. B **62**, 7045 (2000); H. Hu, D.-J. Li, J.-L. Zhu, and J.-J. Xiong, cond-mat/0009044; cond-mat/0010310.
4. A. V. Chaplik, Zh. Éksp. Teor. Fiz. **119**, 193 (2001) [JETP **92**, 169 (2001)].
5. V. Ya. Prinz, V. A. Seleznev, V. A. Samoylov, and A. K. Gutakovsky, Microelectron. Eng. **30**, 439 (1996).
6. H. Ajiki and T. Ando, J. Phys. Soc. Jpn. **62**, 1255 (1993).
7. A. V. Chaplik, Surf. Sci. Rep. **5**, 296 (1985).
8. H. Ajiki and T. Ando, Physica B (Amsterdam) **201**, 349 (1994).
9. M. V. Klein, in *Light Scattering in Solids*, Vol. 3: *Recent Results*, Ed. by M. Cardona (Springer-Verlag, Berlin, 1975; Mir, Moscow, 1986), p. 174.
10. A. O. Govorov and A. V. Chaplik, Zh. Éksp. Teor. Fiz. **98**, 1564 (1990) [Sov. Phys. JETP **71**, 876 (1990)].

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 93, No. 4, 2001, pp. 860–864.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 986–990.

SOLIDS
Electronic Properties

The Doubling of the Anomalous Magnetic Moment of Electron in a Very Strong Constant Homogeneous Electric Field¶

V. I. Ritus

Tamm Department of Theoretical Physics, Lebedev Physical Institute, Russian Academy of Sciences, Moscow, 117924 Russia
e-mail: ritus@lpi.ac.ru

Received April 18, 2001

Abstract—Calculated by the author previously [8], the anomalous magnetic moment (AMM) of the electron in an intense constant electric field changes nonmonotonically as the field increases, passing through a minimum and tending to the doubled Schwinger value for very strong fields. In the present paper, it is supposed that the AMM is related by the Lande factor to the angular momentum of a virtual electron accompanied by a virtual photon. This factor changes its effective value because of the influence of the external field on the motion of the virtual electron and its self-action. With increase of the electric field strength, the virtual electron can successively occupy the excited states $l = 1$, $j = 1/2$ and $l = 1$, $j = 3/2$ in addition to the original state with the orbital angular momentum $l = 0$ and the total angular momentum $j = 1/2$. The first of these excited states decreases the AMM and the second increases and doubles it if only this state is occupied for a very strong field. The latter condition is equivalent to the alignment of the spin and the orbital angular momentum of the electron along the field, while the total angular momentum of the entire system of the virtual electron and the virtual photon remains equal to 1/2. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The purpose of this paper is to draw attention to an interesting dependence of the anomalous magnetic moment of electron on the intensity of the external constant homogeneous electric field. The dependence of the AMM on the constant magnetic or crossed field was considered by Demeur [1], Newton [2], Ternov *et al.* [3], Ritus [4], Jancovici [5], Tsai Wu-yang and Yildiz [6], and Baier *et al.* [7]. It was shown that the AMM tends to zero for a very strong magnetic or crossed field. The dependence of the AMM on the constant electric field cannot be obtained from its dependence on the magnetic field by the analytic continuation $\mathbf{H}^2 \longrightarrow -\mathbf{E}^2$ because of the nonanalyticity at zero field.

In my paper [8], the eigenvalue of the mass operator of electron in a constant homogeneous electromagnetic field of an arbitrary intensity was found. In particular, this eigenvalue, or more precisely, the corresponding elastic scattering amplitude involves the dependence of the AMM $\Delta\mu$ on the electric and magnetic fields. If one keeps only the dependence on the electric field ε and confines oneself to the state with $\mathbf{p}_\perp = 0$, the ratio of $\Delta\mu$ to the Bohr magneton μ_0 is given by

$$\frac{\Delta\mu}{\mu_0} = \frac{\alpha}{\pi} J(\beta), \quad \beta = \frac{\hbar|e\varepsilon|}{m^2c^3}, \tag{1}$$

where β is the electric field ε in characteristic QED units and

$$\begin{aligned} J(\beta) &= 1 - I(\beta), \\ I(\beta) &= \frac{1}{\beta}\int_0^\infty dy \sin(y/\beta)\phi(y), \\ \phi(y) &= \int_y^\infty \frac{dx u^2}{x(1+u)^2}, \\ u &= x[\coth(x-y) - \coth x]. \end{aligned} \tag{2}$$

It was shown that in a weak field ($\beta \ll 1$),

$$\begin{aligned} J(\beta) &= \frac{1}{2} - \frac{4}{3}\beta^2\left(\ln\frac{\gamma}{2\beta} - \frac{23}{12}\right) \\ &- \frac{128}{3}\beta^4\left(\ln\frac{\gamma}{2\beta} - \frac{81}{70}\zeta(3) - \frac{209}{1120}\right) + \ldots \end{aligned} \tag{3}$$

and in a strong field ($\beta \gg 1$),

$$J(\beta) = 1 - \frac{\pi}{4\beta}\left(\ln\frac{2\beta}{\gamma} - 1\right) + \ldots, \tag{4}$$

where $\gamma = 1.781\ldots$ and $\zeta(3) = 1.202\ldots$.

Thus, as the field increases, the AMM first decreases from the Schwinger value $\alpha/2\pi$, reaches a minimum, and then increases and approaches the doubled Schwinger value α/π.

¶This article was submitted by the author in English.

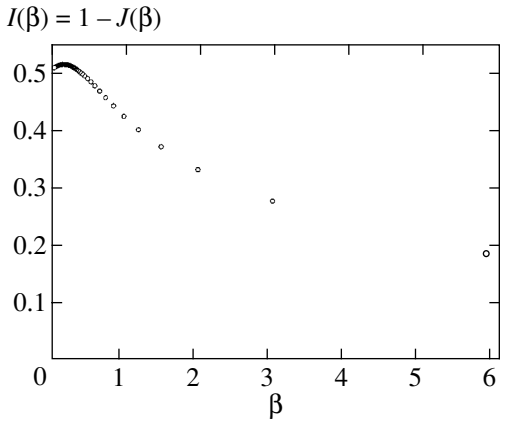

Fig. 1.

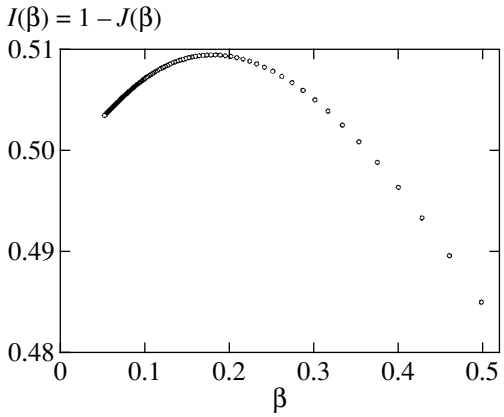

Fig. 2.

This intriguing dependence is also confirmed by the numerical calculation of the integral $J(\beta)$, see Figs. 1 and 2. The minimum of $J(\beta)$ is located at $\beta \approx 0.179$ and is equal to $J(0.179) \approx 0.49040$.

The most striking property of $J(\beta)$ is the doubling of this function and of the AMM in the strong-field limit compared to its value in the weak-field limit,

$$\frac{J(\infty)}{J(0)} = 2. \tag{5}$$

In all other fields (magnetic, crossed) the AMM turns into zero in the strong-field limit.

Another interesting property of $J(\beta)$ is its nonmonotonic dependence with one minimum.

The formula for the AMM is nothing else than the Fourier sine transform of the function $\phi(y)$. This function has a maximum at zero, which is equal to 1/2, and monotonically decreases to zero as $y \longrightarrow \infty$,

$$\begin{aligned}\phi(y) &= \frac{1}{2} - y^2\left(\frac{2}{3}\ln\frac{1}{2y} - \frac{5}{18}\right)\\ &- y^3\left(\frac{1}{3} + \frac{8\pi^2}{45}\right) + \ldots, \quad y \ll 1,\\ \phi(y) &= \frac{1}{2y}(\ln 2y - 1)\\ &- \frac{1}{8y^2}\left(\ln^2 2y - 8\ln 2y + \frac{\pi^2}{3} + 4\right) + \ldots, \quad y \gg 1.\end{aligned} \tag{6}$$

Qualitatively, a similar nonmonotonic behavior for $J(\beta)$ with the minimum at $\beta \sim 1$ and the doubling at infinity, $J(\infty) = 2J(0)$, would be given by the Gaussian

$$\phi(y) = \frac{1}{2}e^{-y^2}$$

and the Lorentzian

$$\phi(y) = \frac{1}{2}(1 + y^2)^{-1}$$

functions of y.

We consider the physical meaning of the function $\phi(y)$ and its argument.

2. THE MASS OPERATOR OF ELECTRON IN A CONSTANT ELECTROMAGNETIC FIELD

The law of motion of a relativistic classical charge in the homogeneous constant electromagnetic field $F_{\alpha\beta}$ can be written as

$$x_\alpha(s) - x_\alpha(0) = \left(\frac{\exp(2eFs) - 1}{eF}\right)_{\alpha\beta}\pi_\beta(0), \tag{7}$$

where $x_\alpha(s)$ is the charge 4-coordinate depending on the proper time s and $x_\alpha(0)$ and $\pi_\alpha(0)$ are the initial 4-coordinate and the kinetic 4-momentum respectively. In a Lorentz system where the electric and magnetic fields are parallel, the electron moves along a helical line with alternating pitch whose rate of change is defined by the electric field and the period of revolution is defined by the magnetic field.

Quantum motion of the electron in an external field, with the radiative corrections taken into account, is described by the Dirac wave equation with the mass operator. Roughly speaking, the mass operator is defined in the e^2-approximation by the product of the causal propagation functions $S^c(x, x')$ and $D^c(x - x')$ of the electron in an external field and of the photon in the vacuum:

$$M(x, x') = ie^2\gamma_\mu S^c(x, x')\gamma_\mu D^c(x - x'). \tag{8}$$

Here and below, we use the same notations as in [8]. In the proper time representation, we have

$$S^c(x, x') = -\frac{ie^{i\varphi}}{16\pi^2}\int_0^\infty \frac{ds e^2 \eta\varepsilon}{\sin(e\eta s)\sinh(e\varepsilon s)}$$

$$\times\left(m - \frac{i}{2}\gamma Bz\right)\exp\left(-im^2 s + \frac{iz\beta z}{4} + \frac{ie\sigma Fs}{2}\right),$$

$$D^c(z) = -\frac{i}{16\pi^2}\int_0^\infty \frac{dt}{t^2}\exp\left(\frac{iz^2}{4t}\right), \quad z = x - x', \tag{9}$$

$$B_{\alpha\beta} = \beta_{\alpha\beta} + eF_{\alpha\beta}, \quad \beta_{\alpha\beta} = (eF\coth(eFs))_{\alpha\beta},$$

$$\varphi = e\int_{x'}^{x} dy_\alpha A_\alpha(y),$$

and ε and η are the strengths of the electric and the magnetic field in the frame where they are parallel. The φ integration goes along the straight line.

After going over to the $E_p(x)$-representation, the mass operator becomes diagonal and its renormalized eigenvalue is given by the γ-matrix,

$$M_R(\bar p, F) = \frac{\alpha}{2\pi}\int_0^\infty\int_0^\infty \frac{ds dt}{t^2}\left\{\frac{\sin(e\eta w_1)\sinh(e\varepsilon w_2)}{\sin(e\eta s)\sinh(e\varepsilon s)}\right.$$

$$\times\exp\left(-im^2 s - i\bar p w \bar p - \frac{ie\sigma Fw}{2}\right)$$

$$\times\left[2m(S + i\gamma_5 P) + i\exp\left(-\frac{ie\sigma Fs}{2}\right)\right. \tag{10}$$

$$\left.\left.\times\gamma\exp(eF(w+s))\frac{\sinh(eFw)}{\sinh(eFs)}\bar p\right]\right.$$

$$\left.-\frac{\omega^2}{s^2}\exp(-im^2 s - ip^2 w)\left(2m + i\gamma\bar p\frac{\omega}{s}\right)\right\} + M_R^0(\bar p).$$

This is Eq. (52) in [8], where one can find all the details about the quantities involved here and the notations.

It is now important to discuss the transformation of the term γBz entering the electron propagator. This term, being linear in the field (for a weak field) and linear in the coordinate difference $z = x - x'$, gives a contribution to the AMM and contains information about the motion of the virtual electron between the points x' and x when it is accompanied by a virtual photon and its motion is distorted by the external field. Because

$$\gamma_\mu\left(m - \frac{i}{2}\gamma Bz\right)\exp\left(\frac{ie\sigma Fs}{2}\right)\gamma_\mu$$

$$= 4m(S + i\gamma_5 P) + i\exp\left(-\frac{ie\sigma Fs}{2}\right)\gamma Bz, \tag{11}$$

the term γBz appears in $M_R(\bar p, F)$ in the second term in the square brackets as

$$\gamma Bz_{\text{eff}} = 2\gamma\exp(eF(w+s))\frac{\sinh(eFw)}{\sinh(eFs)}\bar p$$

$$= \gamma\frac{\exp(eFs)(\exp(2eFw) - 1)}{\sinh(eFs)}\bar p \tag{12}$$

$$= \gamma B\frac{\exp(2eFw) - 1}{eF}\bar p.$$

Therefore, after the integration over x and x' performed in passing from $M(x, x')$ to $M_R(\bar p, F)$, we obtain instead of $z \equiv x - x'$ the quantity

$$z_{\alpha\,\text{eff}} = \left(\frac{\exp(2eFw) - 1}{eF}\right)_{\alpha\beta}\bar p_\beta, \tag{13}$$

which is the "mean" or the "effective" coordinate difference. Here, $\bar p_\beta$ is the constant momentum four-vector that characterizes the quantum motion of the electron in the external field and $w_{\alpha\beta}$ is the 4×4 matrix

$$w = \frac{1}{eF}\operatorname{arccoth}\left(\coth(eFs) + \frac{1}{eFt}\right) \tag{14}$$

with two doubly degenerate eigenvalues

$$w_1(s, t, \eta) = \frac{1}{e\eta}\operatorname{arccot}\left(\cot(e\eta s) + \frac{1}{e\eta t}\right),$$

$$w_2(s, t, \varepsilon) = \frac{1}{e\varepsilon}\operatorname{arccoth}\left(\coth(e\varepsilon s) + \frac{1}{e\varepsilon t}\right) \tag{15}$$

playing the roles of the magnetic and the electric proper times. Thus, the virtual electron moves between the points x' and x of the emission and the absorption of the virtual photon "as a classical charge" with two proper times.

Because the virtual electron is accompanied by a virtual photon with the proper time t (or with the squared momentum $k^2 \sim t^{-1}$), the proper times $w_{1,2}$ are always less than s,

$$0 \le w_{1,2} \le s. \tag{16}$$

The symbol arccot also indicates that w_1 is always in the same period with s,

$$n\pi \le e\eta w_1 \le e\eta s \le (n+1)\pi, \quad n = 0, 1, 2, \ldots.$$

The argument y of $\phi(y)$ is equal to

$$y = e\varepsilon(s - w_2) \equiv e\varepsilon s - \operatorname{arccoth}\left(\coth(e\varepsilon s) + \frac{1}{e\varepsilon t}\right); \quad (17)$$

i.e., it is proportional to the delay of the electric proper time of the virtual electron compared to the proper time of the real electron in the field.

3. PHYSICAL INTERPRETATION OF THE AMM DOUBLING FOR A VERY STRONG ELECTRIC FIELD

The AMM $\Delta\mu$ explicitly appears in the elastic scattering amplitude [8]

$$T(\bar{p}, \bar{s}, F) = -\mathrm{Tr}(M_R(\bar{p}, F)u_{\bar{p}\zeta}\bar{u}_{\bar{p}\zeta}) \quad (18)$$

as the real part of the coefficient at the first of the two spin-dependent invariants

$$\begin{aligned} -\frac{\bar{s}F^*\bar{p}}{m}\mathrm{Tr}(u\bar{u}) &= \bar{u}_{\bar{p}\zeta}\frac{1}{2}\sigma F u_{\bar{p}\zeta}, \\ \frac{\bar{s}F\bar{p}}{m}\mathrm{Tr}(u\bar{u}) &= \bar{u}_{\bar{p}\zeta}\frac{1}{2}\sigma F^* u_{\bar{p}\zeta}. \end{aligned} \quad (19)$$

Here, $\bar{s}$ is the polarization 4-vector, $F^*_{\alpha\beta}$ is the field tensor dual to $F_{\alpha\beta}$, and

$$\begin{aligned} u_{\bar{p}\zeta}\bar{u}_{\bar{p}\zeta} &= \frac{\mathrm{Tr}(u\bar{u})}{4m}(m - i\gamma\bar{p})(1 + i\gamma_5\gamma\bar{s}), \\ \bar{p}^2 &= -m^2, \quad \bar{s}^2 = 1, \quad \bar{s}\bar{p} = 0, \end{aligned} \quad (20)$$

is the polarization density matrix. For the state with $\mathbf{p}_\perp = 0$ and the magnetic field $\eta \longrightarrow 0$, $\Delta\mu$ is given by Eqs. (1) and (2).

The main contribution to the integral $I(\beta)$ comes from $y \sim \beta$. The doubling of the AMM in a strong field is then related to a large delay $y \sim \beta \gg 1$ and the explicit expression for y shows that

$$\begin{aligned} s \sim m^{-2}, \quad t \sim \frac{1}{e\varepsilon} \ll m^{-2} \\ \text{or } k^2 \sim \frac{1}{t} \sim e\varepsilon \gg m^2. \end{aligned} \quad (21)$$

In other words, in a strong field the virtual electron is accompanied by a "heavy" virtual photon with the squared momentum

$$k^2 \sim e\varepsilon \gg m^2.$$

We note that

$$s \sim t \sim m^{-2}$$

in a weak field.

The phenomenological and very speculative explanation of the doubling of the magnetic moment in the system of a virtual electron and a photon with the increase of the virtual photon momentum squared may be as follows.

There is the known relation between the magnetic moment μ and the angular momentum j of an electrodynamical system,

$$\mu = \mu_0 gj, \quad \mu_0 = \frac{\hbar e}{2mc}, \quad (22)$$

where g is the gyromagnetic ratio. Writing this relation for the AMM and comparing it with the definition of the function $J(\beta)$,

$$\frac{\Delta\mu}{\mu_0} = \frac{\alpha}{2\pi}gj, \quad \frac{\Delta\mu}{\mu_0} = \frac{\alpha}{\pi}J(\beta), \quad J(\beta) = \frac{1}{2}gj, \quad (23)$$

shows that $J(\beta)$ can be considered as half the product of the gyromagnetic ratio and the angular momentum of the virtual electron.

In a weak field, $gj/2 = 1/2$ because the virtual electron has the quantum numbers $s = 1/2$, $l = 0$, and $j = 1/2$ and the Lande formula

$$g = 1 + \frac{j(j+1) + s(s+1) - l(l+1)}{2j(j+1)} \quad (24)$$

gives $g = 2$. In a strong field, the virtual electron can go to the state with $s = 1/2$, $l = 1$, and $j = 3/2$, for which $g = 4/3$. Then $gj/2 = 1$ and the AMM doubles.

For moderate field intensities, the virtual electron can be in a superposition of the states $s = 1/2$, $l = 0$, $j = 1/2$ and $s = 1/2$, $l = 1$, $j = 1/2$. Because $g = 2/3$ and $gj/2 = 1/6$ for the latter state, the decrease of the AMM with the increase of β becomes clear until β is sufficiently small and the state with $s = 1/2$, $l = 1$, and $j = 3/2$ is not perceptibly excited.

Thus, the following physical picture can occur.

The electron interacting with itself via a virtual photon possesses the total angular momentum $J = 1/2$, which can be considered as the vector sum of the virtual electron angular momentum $j = 1/2$ and the proper moment (spin) $j_\gamma = 1$ of the virtual photon. The external electric field changes the motion of the virtual electron such that the electron can acquire the orbital angular momentum $l = 1$ and its total angular momentum j can remain equal to 1/2 or become equal to 3/2. Besides, the vector sum

$$\mathbf{J} = \mathbf{j} + \mathbf{j}_\gamma$$

of the virtual electron and the virtual photon angular momenta remains equal to $J = 1/2$ and their projections on the electric field direction satisfy the conservation law

$$m_J = m_s + m_l + m_\gamma. \quad (25)$$

If the orbital angular momentum and the spin of the virtual electron prefer to be parallel as the electric field increases, such that

$$j = l + 1/2$$

and

$$m_j = m_s + m_l = \pm(l + 1/2),$$

then the appearance of the states with $l > 1$ becomes impossible.

ACKNOWLEDGMENTS

The author thanks A.I. Nikishov, A.I. Zelnikov, and A.I. Ritus for fruitful discussions and help. The work was partly supported by the Russian Foundation for the Basic Research project nos. 00-15-96566 and 99-02-17916a.

REFERENCES

1. M. Demeur, Acad. R. Belg., Cl. Sci. **28**, 1643 (1953).
2. R. G. Newton, Phys. Rev. **96**, 523 (1954); Phys. Rev. D **3**, 626 (1971).
3. I. M. Ternov, V. G. Bagrov, V. A. Bordovitsyn, and O. F. Dorofeev, Zh. Éksp. Teor. Fiz. **55**, 2273 (1968) [Sov. Phys. JETP **28**, 1206 (1969)].
4. V. I. Ritus, Zh. Éksp. Teor. Fiz. **57**, 2176 (1969) [Sov. Phys. JETP **30**, 1181 (1970)].
5. B. Jancovici, Phys. Rev. **187**, 2275 (1969).
6. Wu-yang Tsai and A. Yildiz, Phys. Rev. D **8**, 3446 (1973).
7. V. N. Baier, V. M. Katkov, and V. M. Strakhovenko, Yad. Fiz. **24**, 379 (1976) [Sov. J. Nucl. Phys. **24**, 197 (1976)].
8. V. I. Ritus, Zh. Éksp. Teor. Fiz. **75**, 1560 (1978) [Sov. Phys. JETP **48**, 788 (1978)]; Erratum: Zh. Éksp. Teor. Fiz. **76**, 383 (1979) [Sov. Phys. JETP **49**, 194 (1979)].

Journal of Experimental and Theoretical Physics, Vol. 93, No. 4, 2001, pp. 865–875.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 991–1003.

SOLIDS
Electronic Properties

Light-Induced Temporal Broadening of Optical Lines of Individual Molecules in Polymers and Glasses

I. S. Osad'ko

Lebedev Physical Institute, Russian Academy of Sciences, Leninskiĭ pr. 53, Moscow, 117924 Russia
e-mail: osadko@sci.lebedev.ru

Received May 15, 2001

Abstract—The spectral diffusion theory developed for explaining the logarithmic temporal broadening of the optical line of an individual impurity center and based on spontaneous tunnel transitions in a polymer or glass is supplemented by taking into account the tunnel transitions occurring in the center itself upon the absorption of a photon. This light-induced additional tunneling leads to a sharp temporal broadening of the optical line from an individual impurity molecule occurring against the background of a slower logarithmic line broadening. It is shown that the inclusion of light-induced tunneling can explain the difference in the temporal line broadening of three individual Terylene molecules introduced in polyethylene, which was measured in experiments by the group headed by M. Orrit [A. M. Boiron *et al.*, Chem. Phys. **247**, 119 (1999)]. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

According to the standard theory of broadening of optical lines from impurity centers in crystals based on the electron–phonon interaction, the half-width of a phononless line is a function of temperature but does not depend on the measuring time [1]. Consequently, the doubled rate $2/T_2$ of optical dephasing measured in "fast" experiments such as photon echo must coincide with the half-width of the phononless line measured in "slow" experiments such as stable spectral hole burning. Indeed, such a coincidence is observed for impurity centers of crystals; i.e., the phononless line half-width does not depend on the time of measurement [2].

However, a different situation is observed in polymers and glasses. This fact remained unknown for a long time since the strong nonuniform broadening existing in these solvents masks the uniform broadening of the phononless line from an individual molecule. The nonuniform broadening can be eliminated by exposing a polymer with impurity molecules to laser radiation since the light emitted by a laser effectively excites only the impurity molecules whose phononless lines coincide in frequency with the laser emission line. Such a frequency selection of impurity molecules takes place during spectral hole burning in nonuniformly broadened optical bands and during the measurements of a photon echo signal. Consequently, under these conditions, we are dealing with a spectrally homogeneous system in which a uniform phononless line can be hypothetically investigated.

The application of the spectral hole burning method, as well as photon echo to impurity centers in polymers and glasses, made it possible to discover the time dependence of the broadening of the spectral hole [3, 4] burnt in a nonuniformly broadened optical band as well as the time dependence of the optical dephasing rate [5, 6] of impurity centers in polymers and glass. The time dependence of the spectral hole half-width was called the spectral diffusion.

This phenomenon was discovered almost 40 years ago while studying the microwave echo in spin systems [7]. The term spectral diffusion reflects the contemporary level of understanding of this phenomenon, which was interpreted as follows. In the ensemble of spins with coinciding spectral lines (i.e., in a homogeneous ensemble), the line corresponding to different spins are displaced with time by different distances and in different directions on the frequency scale under the action of the spin–spin interaction. This nonuniform broadening increasing with time and resembling the diffuse spreading in space (e.g., of heat from a source) was referred to as spectral diffusion.

When the broadening with time of holes burnt in nonuniformly broadened optical bands of impurity centers of polymers was subsequently discovered [3, 4], the analogy with the spectral diffusion observed earlier in spin systems was noted immediately. The concept of spectral diffusion as a nonuniform broadening in an initially spectrally homogeneous molecular ensemble, increasing with time, was also borrowed. This concept is used, for example, in some of the recent reviews devoted to spectral hole burning in nonuniformly broadened bands of impurity molecules in polymers and glasses [8]. However, the above-described concept of spectral diffusion cannot explain the broadening with time of the optical line of individual molecules, which was observed in recent experiments [9, 10].

1063-7761/01/9304-0865$21.00

It should be noted, however, that another interpretation of spectral diffusion was given as long ago as in the 1960s in the theoretical work by Klauder and Anderson [11]. According to this hypothesis, this phenomenon is the broadening with time of the resonance line of spin due to the dipole magnetic interaction of the resonance spin with a huge number of spins which are not in resonance with the exciting magnetic field. As applied to molecules as two-level systems, such a concept of spectral diffusion presumes that the optical line corresponding even to an individual molecule may broaden with time if this molecule interacts with a vast ensemble of nonequilibrium two-level systems (TLS) of glasses and polymers.

According to the dynamic theory [12, 13], the interaction of a chromophore of an impurity center with a vast ensemble of slowly relaxing TLS existing in abundance in polymers and glasses may lead to broadening in time of the phononless line of each molecule. This radically changes the meaning of the term spectral diffusion as applied to optical spectra since it denotes now the time broadening of the optical line of each molecule. The spectroscopy of individual molecules gives convincing experimental proof of such an interpretation of spectral diffusion.

However, recent measurements of time broadening of optical lines for individual molecules [14] revealed new facts which cannot be explained in the existing theory of spectral diffusion of the dynamic [12, 13] or stochastic [4, 15, 16] types. It should be noted that the above-mentioned theories disregarded the light-induced tunneling in TLS. In this paper, we supplement the dynamic theory of spectral diffusion [12, 13] by the inclusion of light-induced tunneling in TLS and prove that with such a generalization, the dynamic theory of spectral diffusion [12, 13] is able to explain successfully the new experimental facts discovered in [14].

2. FORMULA FOR THE HALF-WIDTH OF THE PHONONLESS LINE OF AN IMPURITY MOLECULE

According to the dynamic theory [1, 12, 13], the half-width of the phononless line corresponding to impurity centers in polymers and glasses can be described as the sum of the following four terms:

$$\Delta\omega_{1/2} = \frac{1}{T_1} + \gamma_{\rm ph}(T) + \gamma_L(T) + \gamma_{\rm TLS}(T, t). \qquad (1)$$

Here, the first term describes the natural half-width of the line, to which the half-width of a phononless line tends as temperature T tends to zero; i.e., T_1 defines the lifetime of the molecule in an excited electronic state. The second term is due to the quadratic Frank–Condon electron–phonon interaction, It is this term that determines the temperature broadening of the phononless line of impurity centers in crystals without TLS [1]. The third and fourth terms appear due to the quadratic Frank–Condon interaction of an optical electron of an impurity molecule with TLS. The third term emerges due to the interaction with several TLS from the local neighborhood of the impurity, while the fourth term is generated by the interaction with a vast number of TLS located at large distances from the given molecule. It is worth noting that the interaction with TLS from the local neighborhood of the impurity does not lead to a time dependence of the half-width of the phononless line. It is only the interaction with a vast number of distant TLS, which is described by the last term, that gives the dependence on time.

This important conclusion of the dynamic theory can easily be illustrated using the formula [12, 13]

$$\gamma_j = \frac{\Delta_j^2 R_j}{\Delta_j^2 + R_j^2}\rho_j(T, t)(1 - \rho_j(T, t)) \qquad (2)$$

describing the contribution of the jth TLS to the phononless line half-width. Here, Δ_j is the parameter characterizing the intensity of quadratic interaction of an optical electron of the impurity molecule with TLS. This parameter determines the variation of splitting of TLS upon the electron excitation of an impurity molecule. Parameter R_j determines the relaxation rate of the TLS, while the parameter

$$\rho_j(T, t) = \rho_j(T, \infty) + (\rho_j(T, 0) - \rho_j(T, \infty))\exp(-R_j t) \qquad (3)$$

defines the probability of finding a TLS in an excited state. Nonequilibrium TLS's must be characterized by relaxation times much longer than the lifetime T_1 of an excited electron level. Formula (2) shows that if, for example, the TLS relaxation occurs on the microsecond scale, the relaxation constant R_j is of the order of 10^6 s^{-1} and, hence, the contribution of such a TLS to the half-width γ_j of a phononless line is of the same order of magnitude; i.e., $\gamma_j \ll 1/T_1$. In accordance with relation (1), its half-width is two orders of magnitude larger. For this reason, a small number of nonequilibrium TLS's (i.e., TLS's with small R_j) make an insignificant contribution to the half-width of the phononless line. It is only TLS's with $R_j > 1/T_1$ that make a noticeable contribution to the half-width of this line. However, such TLS's relax rapidly, i.e., can be regarded as equilibrium systems like phonons. Only the interaction of a chromophore with a vast number of nonequilibrium TLS's, which is described by the term $\gamma_{\rm TLS}(T, t)$ in formula (1), may lead to a noticeable time dependence of the phononless line half-width since the small contribution to the half-width from each TLS is compensated by the huge number of such TLS's. Obviously, the contribution to the half-width of the line from nonequilibrium TLS's from the local neighborhood of the impurity can be neglected. This means that the time broadening of the optical line of different molecules must be the same.

The form of the time dependence of the half-width $\gamma_{TLS}(T, t)$ is determined by the type of electrostatic interaction between an impurity and TLS's [17, 18]. For the dipole–dipole type of this interaction, the half-width increases with time according to a logarithmic law; i.e.,

$$\gamma_{TLS}(T, t) = aT\log R_m t, \quad (4)$$

where R_m is the maximum relaxation rate for TLS's, whose order of magnitude is estimated as 10^9–10^{10} s^{-1}. Since this term is due to the interaction with a vast number of TLS's separated from a given impurity by considerable distances, it must be practically the same for all impurity molecules. However, Fig. 1 demonstrates, on the contrary, different time dependences of the half-width for different individual molecules. In the following sections, we will prove that this contradiction can be eliminated by taking into account the fact that formula (1) for the half-width takes into account only spontaneous relaxation of TLS's. If, however, we also take into account the effect of light-induced transition in TLS's on the half-width of the line, the experimental results presented in Fig. 1 can be interpreted successfully.

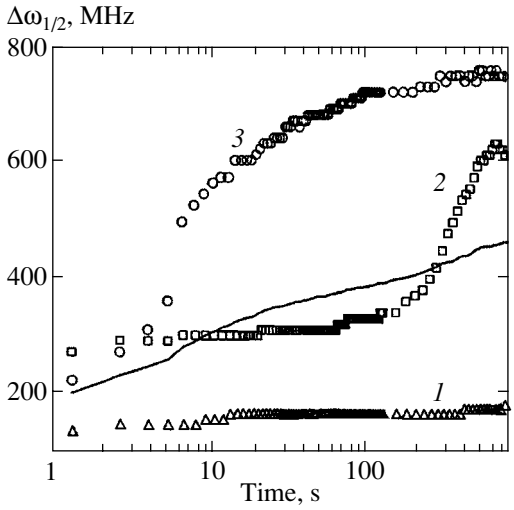

Fig. 1. Time variation of half-widths of the optical lines of three individual Terylene molecules in polyethylene and the time dependence of the half-width of the line corresponding to an ensemble of eight molecules (solid curve) [14].

3. MANIFESTATION OF TIME BROADENING OF THE LINE IN A TWO-PHOTON CORRELATOR

In experiments with individual molecules, the photons emitted by a molecule have to be counted. The profile of the optical line of an individual molecule can be measured using single- or two-photon counting methods. When the single-photon counting method is used, all the photons emitted by an individual molecule are counted; i.e., the intensity of fluorescence is measured. In two-photon methods, photon pairs with a time delay between the photons of a pair are counted. It will be shown below that both counting methods provide the same information on the dynamics of an individual impurity center.

In single-photon counting methods, the results of measurements, as a rule, is a random quantity. This can be, for example, the time dependence of the intensity of emitted light or of the frequency of absorbed light, i.e., the "spectral trajectory." The spectral trajectory is a random function which is not reproduced in experiments. It cannot be calculated theoretically since quantum mechanics makes it possible to calculate only the probability of observing a certain frequency value. The information on probabilities is extracted from the measured spectral trajectory by its statistical processing. These probabilities can be compared with calculated values. In contrast to single-photon methods, two-photon methods make it possible, on the contrary, to measure probabilities directly in the real time of experiments. For this reason, we will first find out how the temporal evolution of the optical line width, which is a probabilistic parameter, can be measured using two-photon counting methods.

A molecule exposed to light emitted by a CW laser emits fluorescence photons at random instants of time. The sequence of emitted photons reaches a photodetector counting the number of photon pairs with a certain time delay t_0 between the photons in a pair. The counting rate $p(t_0)$ for such pairs was referred to in [19, 20] as a two-photon correlator. Such a correlator can be calculated if the Hamiltonian of the system is known. The quantum theory of two-photon correlators was constructed in [13, 19]. According to this theory, a two-photon correlator is defined by the simple formula

$$p = \frac{n_1(t_0, \Delta, T)}{T_1}, \quad (5)$$

where n_1 is the probability of detecting an impurity center in the fluorescent state provided that upon continuous laser pumping, it is in the ground state with the probability equal to unity at instant $t_0 = 0$. It was shown in [13, 19] that a two-photon correlator is a function of the delay time t_0 between the photons in a pair, of the difference Δ between the lasing frequency and the frequency corresponding to the peak of the phononless line, and of temperature; i.e., $p(t_0, \Delta, T)$.

In [20], the correlator was calculated for a simple case (typical of organic molecules) when the ground and the first excited singlet levels are separated by a triplet layer. The correlator of such a molecule is shown in Fig. 2. Obviously, the molecule subjected to continuous laser pumping performs multiple "jumps" at random instants between the ground and excited singlet states, getting into the triplet state from time to time. As long as the molecule is in the triplet state, fluorescence

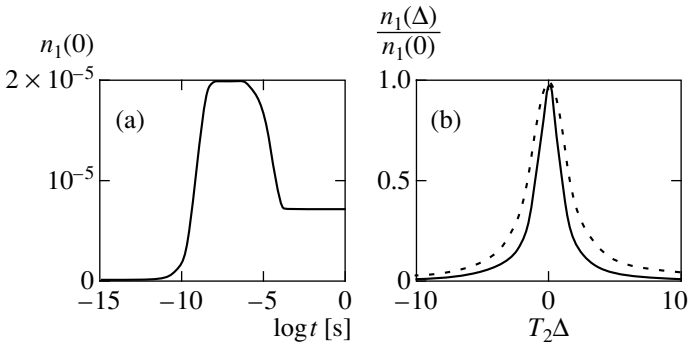

Fig. 2. Two-photon correlator of an individual molecule with a triplet level as a function (a) of the time delay between the photons of a pair and (b) of the frequency of pumping radiation for $t_0 = 10^{-6}$ (solid curve) and 10^{-4} s (dashed curve).

photons are not emitted. For this reason, the train of photons is interrupted intermittently. The photons will be as if grouped on the time scale (photon bunching) [21]. Such a scintillating fluorescence is typical of nanoobjects. For example, it can also be observed during the detection of radiation emitted from quantum dots of semiconductors [22, 23] and in the study of the kinetics of chemical reactions at individual molecules with the help of fluorescence methods [24, 25]. These time intervals with and without fluorescence are usually referred to as *on* and *off* intervals, respectively.

The correlator as a function of frequency describes the shape of the absorption line. Consequently, calculating such a correlator for different delay times between photons, we can obtain information on the temporal behavior of the absorption line of an impurity center. Figure 2 shows the time and frequency dependences of a two-photon correlator calculated in [19, 20] for a molecule with a triplet level. The exponential increase or exponential decay on the logarithmic time scale used in Fig. 2a appears as a smooth step extended over an order on the time scale. For time intervals of the order of 10^{-4} s, a smooth step describes the exponential decrease in the probability of finding the molecule in the excited singlet (i.e., fluorescent) state. Figure 2b shows that the increase in the width of the correlator, which is regarded as a function of the frequency of pumping radiation, also occurs at the same time interval of the order of 10^{-4} s. This broadening of the absorption line has a simple physical meaning: it indicates the broadening of a hole in the probability of finding the molecule in the ground state at the given frequency of the pumping laser radiation due to the increase in the probability of finding the same molecule in the triplet (*off*) state. The same effect in the spectroscopy of holes is known as the effect of triplet saturation of spectral hole [26].The optical line broadening due to the "leakage" of probability to the nonfluorescent triplet *off* state, occurring on a segment of the time scale equal to an order of magnitude of the quantity, resembles the broadening presented in Fig. 1. Since such a broadening of the absorption line occurs after the transition of the molecule from the excited siglet state to the triplet state, it is determined by the light intensity, i.e., is a light-induced broadening. However, the broadening in this case is governed by the intramolecular mechanism of intercombination conversion. Consequently, it must be identical for all impurity molecules, which contradicts Fig. 1. However, the above arguments show that we will be able to explain the experimental results presented in Fig. 1 if we find the *off* states of another physical origin than the triplet state of a molecule. These states must reflect the peculiarities of the local surroundings of an individual molecule.

4. QUALITATIVE INTERPRETATION OF THE LIGHT-INDUCED LINE BROADENING FOR AN INDIVIDUAL MOLECULE

Since individual impurity molecules in a polymer have different surroundings, it is natural to assume that different time broadening of the lines of individual molecules is associated with the interaction of a chromophore impurity molecule with TLS's from the nearest neighborhood of the molecule. The energy level diagram for a chromophore molecule interacting with a TLS is shown in Fig. 3. The meaning of all the constants determining the dynamics of the TLS is clear from the figure. The dynamic theory of the time broadening of optical lines, developed in [12, 13], takes into account the interaction of the chromophore impurity with a vast number of TLS's of the polymer, but only the tunnel transitions between the wells in the ground electron state of chromophore were included; i.e., it was assumed that $B = b = 0$. In other words, this theory disregarded light-induced jumps in TLS's. It will be proved in the next section that disregard for light-induced transitions is justified if we investigate the emergence of the time dependence in the term $\gamma_{TLS}(T, t)$ taking into account the effect of a vast number of TLS's on the line half-width. However, the curves in Fig. 1 cannot be explained if we disregard the light-induced tunneling in TLS's.

Let us first try to explain qualitatively why the line of a molecule can be broadened upon an increase in the time of measurements, which is demonstrated in Fig. 1. For this purpose, we must first describe how the experimental results were obtained in [14]. An individual molecule exposed to light emitted by a CW laser "hopped" between the ground and excited electron states, being multiply excited and emitting each time a fluorescence photon. The sequence of all photons emitted by the molecule was registered; i.e., the single-photon counting method was used. The frequency of laser radiation was scanned so that a frequency interval of several gigahertz was covered in 1 s. When the frequency of laser radiation coincided with the phononless line of the molecule, the latter emitted fluorescence photons which served as a source of information on the light absorption by the molecule. In the experiment, two-dimensional patterns depicted in Fig. 4 were obtained. Dark vertical bands with gaps (trails) visualize the light absorption by an individual molecule at a given instant at a given frequency. It can be seen that the molecules hop at random instants between two spectral positions. In accordance with Fig. 4, about 1000 laser scans were made.

The half-widths of the lines presented in Fig. 1 were calculated in [14] after double averaging. The first averaging was carried out over n laser scans. The time t plotted along the abscissa axis in Fig. 1 is equal to nt_{sc}, where t_{sc} is the duration of a scan. The shape of the spectral line plotted as a result of such an averaging was subjected to noticeable fluctuations (especially for small values of t); for this reason, it was averaged over the entire length of a trail. The half-width of such an averaged line is plotted in Fig. 1.

We assume that the impurity center under investigation, by which we mean an impurity molecule with its nearest neighbors, can be transformed with a change in the atomic arrangement; i.e., the atoms of the impurity center have a couple of equilibrium positions. In this case, the diagram of its electron levels corresponds to that presented in Fig. 3. Such an impurity center has two spectral lines. Let us suppose that this impurity center is excited at the frequency of the 0–1 transition, while the frequency of the 2–3 transition lies outside the frequency interval of the laser scan. In other words, we consider the situation when, for example, only the left trail from the two shown in Fig. 4 is within the laser scan. In this case, the following pattern is observed in each laser scan lasting 1 s. When the lasing frequency coincides with the optical line of the 0–1 transition, the photomultiplier detects a sequence of fluorescence photons with random intervals. Each fluorescence photon is emitted during the transition of the molecule from state 1 to state 0. Consequently, the left pair of wells in Fig. 3 describes the *on* state of the chromophore. However, the system can hop to the right pair of wells owing to the possible tunneling between the left and right pairs of wells. Being in this conformation state, the impurity center has another resonance frequency lying outside the interval in which the frequency of laser radiation is being scanned; i.e., it is not excited by the laser and, hence, does not emit photons. Consequently, the right pair of wells in Fig. 3 corresponds to the *off* state. After

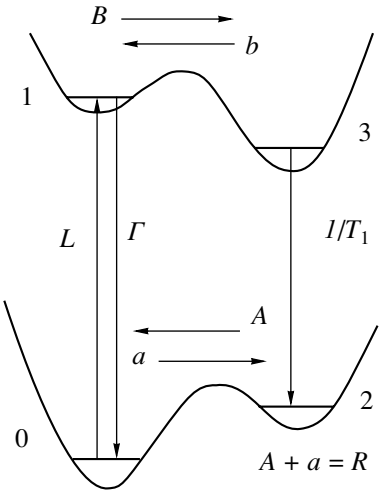

Fig. 3. Energy level diagram for a chromophore molecule interacting with a TLS.

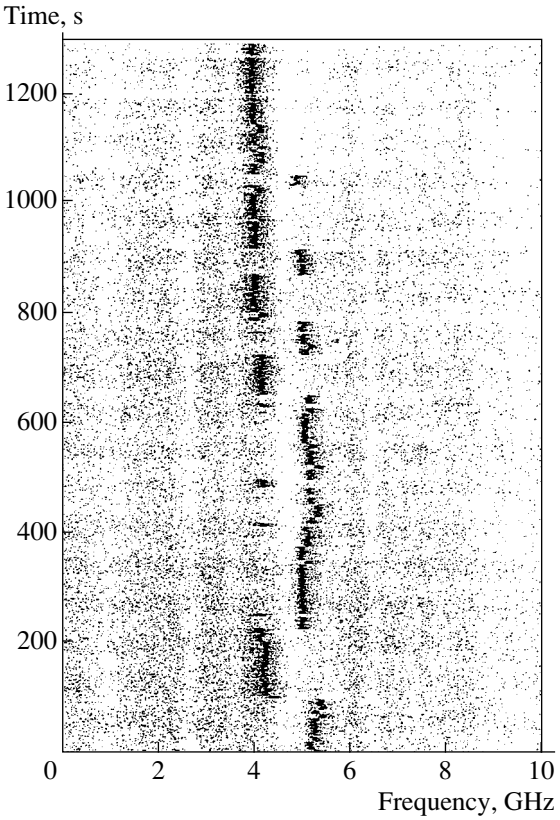

Fig. 4. Trails obtained as a result of scanning of the pumping laser frequency, visualizing the light absorption by a Terylene molecule in polyethylene [14].

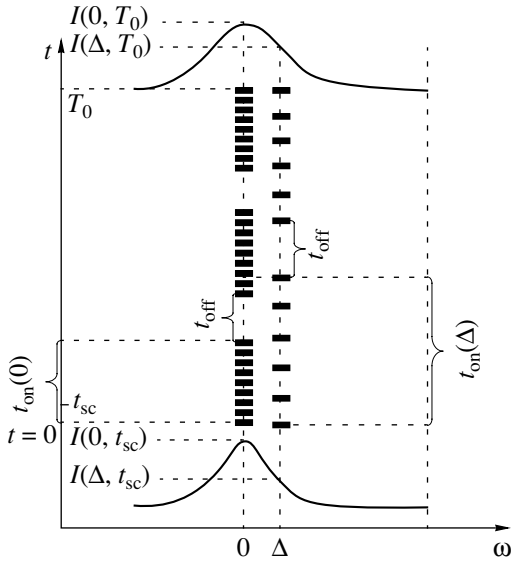

Fig. 5. Schematic diagram explaining the time broadening of the line corresponding to an individual molecule due to light-induced tunneling.

a certain time approximately equal to the reciprocal tunneling probability $1/A$ from state 2 to state 1, the impurity center can again be found in the state described by the left pair of wells and, hence, starts emitting a new sequence of photons in each laser scan.

Figure 5 shows schematically the sequence of photons emitted by the molecule under continuous excitation to the peak of the absorption line of the molecule and to the wing of this line. The sequence of emitted photons splits into groups separated by gaps (photon bunching). In Fig. 5, average values of *on* and *off* intervals are presented. In a real system, the duration of *on* and *off* intervals fluctuates (see Fig. 4). The replacement of fluctuating intervals by their mean values does not change the essence of the explanation given below.

Obviously, a displacement of the lasing line by Δ from the resonance reduces the absorption of light by an impurity center. For this reason, the intervals between fluorescence photons in a group are larger in the case of nonresonant excitation (see Fig. 5). Consequently, the photon count rate $dN(0, t)/dt$ in the case of excitation corresponding to the resonance state is higher than the count rate in the case of excitation to the wing of the line. Besides, if the tunneling in the excited electron state plays the leading role in the transition of the system from the left pair of wells to the right pair, the rate of such a light-induced tunneling is proportional to BL/Γ, where the ratio L/Γ determines the occupancy of level 1. This means that the weaker the light absorption at the frequency of the 0–1 transition, i.e., the smaller the value of L, the lower the rate of departure of the impurity center from the optically active *on* state to the optically inactive *off* state. Consequently, *on* intervals $\tau_{on}(\Delta)$ for the excitation corresponding to the wing of the optical line are longer than the intervals $\tau_{on}(0)$ for the excitation to the resonance state. On the other hand, the rate of departure from the *off* state corresponding to the right pair of wells is determined by constant A; i.e., this rate does not depend on the intensity of absorption. For this reason, the duration τ_{off} of an *off* interval is independent of whether it is excited at a frequency corresponding to the peak of the line or to its wing. Figure 5 reflects all these details of broadening.

If the time $t = nt_{sc}$ of measurement is much shorter than the lifetime of a molecule in the *on* state, i.e., $t \ll \tau_{on}$, the photons being counted belong to one of the groups presented in Fig. 5. In this case, the signal at the photomultiplier is proportional to the rate of emission of photons by the molecule, and we can write the following expression for the ratio of signals in the case of resonant and nonresonant excitation:

$$\frac{I(0, t)}{I(\Delta, t)} = \frac{dN(0, t)/dt}{dN(\Delta, t)/dt}. \tag{6}$$

When the signal is accumulated over a long period of time $T_0 \gg \tau_{on}, \tau_{off}$ depicted in Fig. 5, we have the following ratio for signals:

$$\frac{I(0, T_0)}{I(\Delta, T_0)} = \frac{3\tau_{on}(0)dN(0, t)/dt}{2\tau_{on}(\Delta)dN(\Delta, t)/dt} < \frac{I(0, t)}{I(\Delta, t)}. \tag{7}$$

Indeed, it follows from Fig. 5 that $3\tau_{on}(0) < 2\tau_{on}(\Delta)$. Formula (7) indicates that the half-width of the optical absorption line, measured during time t shorter than τ_{on}, is smaller than the half-width measured during the time T_0 exceeding the time τ_{on}. Precisely this pattern is depicted in Fig. 1.

5. THE THEORY OF LIGHT-INDUCED BROADENING OF OPTICAL LINE

Let us now derive the formula for describing the experimental results on emission line broadening for three individual molecules presented in Fig. 1. We consider an impurity molecule interacting with a huge number of TLS's in a polymer. We assume that only one TLS is present in the local neighborhood; i.e., such an impurity center is characterized by only two optical lines (see the energy level diagram in Fig. 3).

We analyze the case when the pumping by light is carried out only at the frequency of the 0–1 transition. In this case, the system of balance equations with transitions presented in Fig. 3 has the form

$$\begin{aligned}
\dot{\rho}_1 &= -(\Gamma + B)\rho_1 + L\rho_0 + b\rho_3, \\
\dot{\rho}_3 &= B\rho_1 - \left(\frac{1}{T_1} + b\right)\rho_3, \\
\dot{\rho}_0 &= \Gamma\rho_1 - (L + a)\rho_0 + A\rho_2, \\
\dot{\rho}_2 &= a\rho_0 + \frac{\rho_3}{T_1} - A\rho_2,
\end{aligned} \tag{8}$$

where $L = k_{0-1}$ and $\Gamma = L + 1/T_1$. We assume that the constants describing the rates of transitions satisfy the following inequalities:

$$\Gamma \gg L \gg A, B > a, b; \tag{9}$$

which is usually observed in real systems. With such a relation between the rate constants, the time evolution of probabilities ρ_j has two stages (short and long) reflecting the rapid and slow evolution of probabilities ρ_j. The rapid evolution occurs over times of the order of T_1. This evolution corresponds to an increase in the correlator, for example, the same as in Fig. 2a. However, we will be interested in the slow relaxation observed after the stabilization of quasi-equilibrium between the populations ρ_1 and ρ_3 of the electronically excited states, on the one hand, and the population ρ_0, on the other hand. The relation between the probabilities in the quasi-equilibrium state can be determined by putting $\dot{\rho}_1 = \dot{\rho}_3 = 0$. In this case, taking into account inequalities (9), we obtain the following relations from the first and second equations:

$$\begin{aligned}\rho_1 &= \frac{(1/T_1 + b)L\rho_0}{(\Gamma + B)/T_1 + b\Gamma} \approx \frac{L}{\Gamma}\rho_0,\\ \rho_3 &= \frac{BL\rho_0}{(\Gamma + B)/T_1 + b\Gamma} \approx \frac{L}{\Gamma}T_1 B\rho_0.\end{aligned} \tag{10}$$

Substituting relations (10) into the last two equations in system (8), we arrive at the following system of equations:

$$\begin{aligned}\dot{\rho}_0 &= -(\tilde{B} + a)\rho_0 + A\rho_2,\\ \dot{\rho}_2 &= (\tilde{B} + a)\rho_0 - A\rho_2,\end{aligned} \tag{11}$$

describing the slow relaxation in an impurity center with TLS. Using relation $\rho_2(t) = 1 - \rho_0(t)$, we can easily obtain the following equations from Eqs. (10):

$$\dot{\rho}_0 = -(\tilde{B} + R)\rho_0 + A, \tag{12}$$

which includes a function of frequency $\tilde{B}$ determined by the intensity of laser radiation. The solution of Eq. (12) has the form

$$\begin{aligned}\rho_0(t) &= \frac{A}{\tilde{B} + R}\\ &+ \left(\rho_0(0) - \frac{A}{\tilde{B} + R}\right)\exp[-(\tilde{B} + R)t].\end{aligned} \tag{13}$$

Here

$$\tilde{B} = \frac{L}{\Gamma}B, \quad R = A + a. \tag{14}$$

In accordance with formulas (10) and (13), the slow relaxation of all populations is determined by the tunneling between states 0 and 2 as well as between 1 and 3. The transition between states 1 and 3 is light-induced tunneling. Its rate $\tilde{B}$ is proportional to pumping L.

It was noted above that polymers and glasses contain a vast number of TLS's inherent in the solvent, which exist irrespective of the presence of an impurity in the polymer. The number of such TLS's is huge and may considerably exceed the number of impurity centers. The rates of spontaneous tunneling, b and B, in such TLS's with an excited chromophore are comparable with the tunneling rates A and a in the ground electronic state of the chromophore. When such TLS's are taken into account, the light-induced transitions between states 1 and 3 can be disregarded in accordance with formulas (10) and (14), since $L/\Gamma \ll 1$. Such a situation is realized for a chromophore interacting with TLS's in polymers and glasses, and it is the contribution from these TLS's to the half-width $\gamma_{TLS}(T, t)$ that leads to the time dependence of the half-width of the line.

However, the probability of light-induced transitions may exceed the probability of spontaneous transitions. This is observed for TLS's simulating an impurity center. It is these TLS's that participate in photochemical burning of stable spectral holes in nonuniformly broadened optical bands. In the spectroscopy of individual molecules, burning appears as the disappearance of the optical line of a molecule in the spectral range under investigation for a considerable time interval. As a rule, such TLS's describe the state of the chromophore itself; i.e., they belong to the type of TLS's introduced by an impurity molecule into the solvent. Obviously, the number of such TLS's is approximately equal to the number of impurity centers. Experiments show that in this case, a stable spectral hole is burnt in a nonuniformly broadened optical band precisely due to light-induced tunneling. Since this hole exists for several days and even weeks in the absence of light pumping at low temperatures, this indicates an extremely low intensity for such a tunneling in the ground electronic state, i.e., the smallness of constants A and a in the TLS associated with the impurity. Consequently, light-induced tunneling cannot be neglected for such TLS's since it plays a decisive role for explaining the shape of curves *2* and *3* in Fig. 1.

The absorption line obtained as a result of n laser scans obviously reflects the dependence of the probability of observing a fluorescence photon on the frequency of pumping laser radiation. The probability of the emission of a photon is proportional to the probability of finding a chromophore in an excited state, i.e., the probability $\rho_1(t)$, where $t = nt_{sc}$. Using formulas (10) and (13), we obtain

$$\begin{aligned}\rho_1(\Delta, t) &= \frac{L}{\Gamma}\\ &\times\left[\frac{A}{\tilde{B} + R} + \left(1 - \frac{A}{\tilde{B} + R}\right)\exp[-(\tilde{B} + R)t]\right],\end{aligned} \tag{15}$$

where the initial condition $\rho_0(0) = 1$ has been used. Substituting the value of $\tilde{B}$ defined by Eq. (14) into this expression and taking into account the fact that

$$L = 2\chi^2 \frac{1/T_2}{\Delta^2 + 1/T_2^2}, \tag{16}$$

where $\chi = dE/\hbar$ is the Rabi frequency, we arrive at the following expressions for the population for short and long observation times:

$$\rho_1(\Delta, t) = \begin{cases} \dfrac{L}{\Gamma} \approx 2\chi^2 \dfrac{T_1/T_2}{\Delta^2 + 1/T_2^2}, & t \approx 0 \\ 2\chi^2 \dfrac{T_1/T_2}{\Delta^2 + 1/T_2^2(1 + 2\chi^2 T_2 T_1 B/R)}, & t \gg \dfrac{1}{\tilde{B} + R}, \end{cases} \tag{17}$$

the transition from the Lorentzian with half-width $2/T_2$ to the Lorentzian with the half-width

$$\Delta\omega_{1/2}(\infty) = \frac{2}{T_2}\sqrt{1 + \frac{2\chi^2 T_2 T_1 B}{R}} \tag{18}$$

occurring for $t \propto 1/(\tilde{B} + R)$.

6 COMPARISON OF THEORY WITH EXPERIMENT

The spectral function $F(\Delta, t)$ measured with the help of laser frequency scanning during the time $t = nt_{sc}$ is proportional to the function $\rho_1(\Delta, t)$ described by formula (15); i.e.,

$$\rho_1(\Delta, t) = 2\chi^2 T_1 T_2 F(\Delta, t). \tag{19}$$

If we disregard light-induced transitions by equating constant B in formula (15) to zero, the upper and lower expressions in (17) coincide; i.e., the half-width of the spectral function $\Delta\omega_{1/2}$ is equal to $2/T_2$ and may depend on time only due to the time dependence of the half-width $\gamma_{TLS}(T, t)$ appearing in $2/T_2$, which means that it is the same for all molecules. In this case, formula (1) can be transformed to

$$\Delta\omega_{1/2} = \frac{2}{T_2(t)} = \frac{2}{T_2(1)}(1 + C\log t). \tag{20}$$

Coefficients $1/T_2(1)$ and C may be different for different individual impurity molecules in a polymer. Formula (20) correctly describes the time broadening of the line corresponding to molecule 1 in Fig. 1 for $1/T_2(1) = 1.9 \times 10^9$ s^{-1} and $C = 0.08$. This means that molecule 1 does not interact with a TLS characterized by a noticeable probability of light-induced tunneling. Obviously, the line broadening will be correctly described by formulas (1) and (4) in the time interval under investigation even in the case when the impurity center still interacts with the TLS, but the latter has a very large relaxation constant R in the ground state. In this case, in accordance with (18), we have $\Delta\omega_{1/2}(\infty) = 2/T_2$.

Let us now consider the time broadening of the optical line of molecule 2, presented in Fig. 1. It cannot be described by formulas (1) and (4) only. The optical line half-width for this impurity center must be calculated using formulas (15)–(18). In the presence of an optically inactive state, the sequence of photons is interrupted by intermissions; i.e., *on* and *off* intervals exist. Let us find the mean values and durations of *on* and *off* intervals.

Obviously, as long as the molecule hops between states 0 and 1 presented in Fig. 3, the emitting system is in the *on* state, i.e.,

$$\rho_{on} = \rho_0 + \rho_1 \tag{21}$$

is the probability of finding the system in the *on* state. Assuming that the system is in the *on* state with the probability equal to unity, we will find the average lifetime τ_{on} in this state. The kinetics of departure of the system from the *on* state is determined by the following equations:

$$\begin{aligned} \dot{\rho}_1 &= -(\Gamma + B)\rho_1 + L\rho_0, \\ \dot{\rho}_0 &= \Gamma\rho_1 - (L + a)\rho_0. \end{aligned} \tag{22}$$

Summing up these equations, we obtain

$$\dot{\rho}_{on} = -B\rho_1 - a\rho_0. \tag{23}$$

Taking into account inequalities (9), we find (in the same way as in Section 5) that

$$\rho_1 \approx \frac{L}{\Gamma}\rho_0. \tag{24}$$

Using formulas (21) and (24), we can transform Eq. (23) to the following equation:

$$\dot{\rho}_{on} = -\frac{\rho_{on}}{\tau_{on}}, \tag{25}$$

where τ_{on} defined by the formula

$$\frac{1}{\tau_{on}(\Delta)} = \frac{BL + a\Gamma}{L + \Gamma} \tag{26}$$

is the lifetime of the system in the *on* state. It depends on the difference between the laser frequency and the peak of the phononless line. Equation (25) leads to the following expression for the density of the probability of finding an *on* interval of duration t:

$$w_{on}(t) = \frac{1}{\tau_{on}}\exp\left(-\frac{t}{\tau_{on}}\right). \tag{27}$$

Here, τ_{on} is defined by formula (26). Using this probability density, we arrive at the following expression for the mean value of *on* intervals:

$$\int_0^\infty t w_{on}(t)dt = \tau_{on}. \quad (28)$$

Obviously, after performing a quantum jump to the *off* state, the system enters state 2 over a time of the order of a few nanoseconds. Consequently, we can put

$$\rho_{off} = \rho_2. \quad (29)$$

The kinetics of the system departure from the *off* state is defined by the equation

$$\dot{\rho}_{off} = -A\rho_{off}. \quad (30)$$

Consequently, the lifetime of the *off* state and, hence, the mean value of the *off* interval are defined by the formula

$$\tau_{off} = 1/A \approx 1/R. \quad (31)$$

Let us consider formulas (26) and (31) describing the mean values of the *on* and *off* intervals. Formula (26) implies that for $B = a$, the mean duration of the *on* intervals does not depend on laser pumping and is equal to $\tau_{on} = 1/a$. In this case, in accordance with Fig. 5, the duration of the emission of a group of photons in the right column is the same as in the left column. Consequently, the Lorentzian corresponding to the upper part of Fig. 5 and measured during the time T_0 has the same half-width as the lower one measured during a shorter time. Thus, the results presented in Fig. 1 cannot be interpreted for $B = a$. However, the energy level diagram presented in Fig. 3 shows that the probabilities B and a determine the rates of transitions with the creation and annihilation of phonons, respectively. Consequently, these quantities do not coincide since they have different temperature dependences. Their coincidence for a certain temperature may only be accidental. This means that the equality $B = a$ is possible only for a different energy level diagram, namely, if the left well in the upper pair of the potential wells in Fig. 3 is lower on the energy scale than the right well.

Let us return to the energy diagram in Fig. 3. In this case, $B \gg a$. This inequality obviously holds at a low temperature since a tends to zero upon a decrease in temperature, while B remains constant. In this realistic case, formula (26) assumes the form

$$\frac{1}{\tau_{on}(\Delta)} \approx B\frac{L}{\Gamma} \approx BLT_1; \quad (32)$$

i.e., the mean duration of an *on* interval is inversely proportional to laser pumping. This feature readily determines the light-induced nature of the transition from the *on* to the *off* state. Taking into account the fact that

$$\frac{1}{\tau_{on}(0)} = 2\chi^2 T_2 T_1 B, \quad \frac{1}{\tau_{off}} \approx R, \quad (33)$$

we can write formula (18) in the form

$$\Delta\omega_{1/2}(\infty) = \frac{2}{T_2}\sqrt{1 + \frac{\tau_{off}}{\tau_{on}(0)}}. \quad (34)$$

It is also clear that the time interval over which the half-width of the optical line increases is defined by the formula

$$\frac{1}{\tau} = \tilde{B} + R \approx \frac{1}{\tau_{on}(\Delta)} + \frac{1}{\tau_{off}}. \quad (35)$$

The last two formulas are convenient since they do not contain unknown parameters of the type of tunnel transition probabilities. All the quantities appearing in these formulas can be measured comparatively easily in a single experiment. Unfortunately, the mean intervals τ_{on} and τ_{off} were not measured in [14] and we had to select the parameters τ and $r = \tau_{off}/\tau_{on}$. The first parameter determines the time at which the broadening in Fig. 1 starts increasing. The second parameter defines a sort of amplitude of half-width variation. Knowing these parameters, we can calculate the relaxation constants B and R from formulas (33).

The matching of the theoretical curve with the experimental data for half-width in Fig. 6b was attained for the following values of parameters: $r = 3.7$ and $\tau = 234$ s. This corresponds to $\tau_{on} = 297$ s and $\tau_{off} = 1100$ s. Using these values, we can find the tunneling constants B and R for the impurity center 2. The reciprocal value of τ_{off} is the required rate R of tunneling in the ground electronic state. In order to find constant B, we must know the Rabi frequency. It can be expressed in terms of the wavelength λ of absorbed light and the radiation intensity I of the pumping laser expressed in photon/(cm^2 s):

$$\chi^2 = \frac{3}{4\pi}\frac{\lambda^2}{T_1}I. \quad (36)$$

In [14], a laser with a luminous intensity of 1 mW/cm^2 generating a photon flux $I \approx 2.5 \times 10^{15}$ photon/(cm^2 s) was used. Substituting the latter value into formula (36) and taking $\lambda = 500$ nm and $T_1 = 3 \times 10^{-9}$ s [27], we obtain $\chi \approx 2.2 \times 10^7$ s^{-1}. Using curve *2* in Fig. 1, we can determine parameter T_2. Calculations give $1/\pi T_2 = 280$ MHz; i.e., $T_2 \approx 10^{-9}$ s. Substituting the value of the Rabi frequency obtained above and the values of phase-relaxation constant T_2 into formula (33), we obtain $B = 1.1$ s^{-1}.

Figure 6a shows the variation of the population of the fluorescent state 1. The sharp decrease of the curve for $\log(t) > 1.5$ corresponds to the transition of the mol-

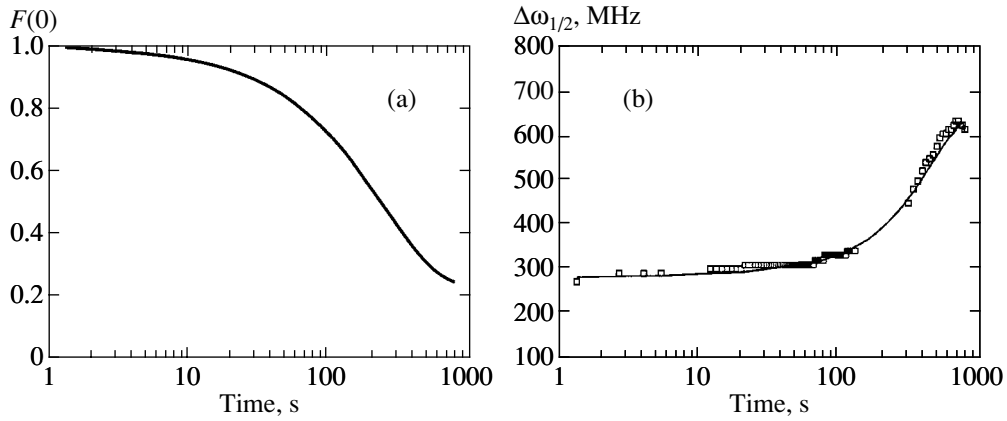

Fig. 6. (a) Time dependence of the population of the fluorescent state of molecule 2 and (b) comparison of the measured line half-width of molecule 2 (symbols) with that calculated by formulas (15) and (20) for $1/\pi T_2 = 280$ MHz, $C = 0.01$, $r = \tau_{off}/\tau_{on} = 3.7$, and $\tau = 234$ s.

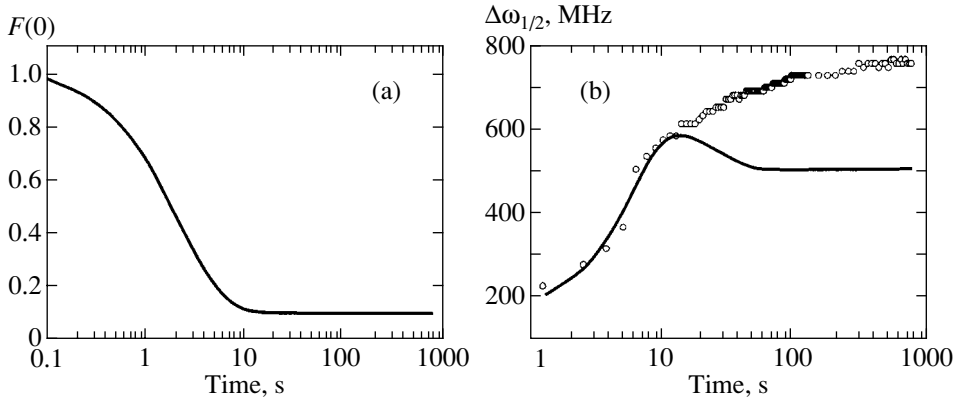

Fig. 7. The same as in Fig. 6 for molecule 3 for $1/\pi T_2 = 238$ MHz, $C = 0.02$, $r = \tau_{off}/\tau_{on} = 9.2$, and $\tau = 2.35$ s.

ecule to the *off* state. A comparison of Figs. 6a and 6b reveals the clearly manifested correlation between the time during which the transition from level 1 to the *off* state occurs and the time during which the half-width of line *2* increase sharply.

A more complicated situation is encountered in the analysis of the line broadening for molecule *3* in Fig. 1. First, it increases rapidly and then at a lower rate over a time interval equal approximately to two orders of magnitude. Since the exponential increase lasts only during a time interval of one order of magnitude, it can be expected that curve *3* cannot be described correctly with the help of formula (15) containing only one exponential. Indeed, Fig. 7b shows that the theoretical curve satisfactorily describes only the initial segment of time broadening for molecule *3*, which occupies only one order on the time scale. Obviously, an auxiliary mechanism of light-induced tunneling also exists and manifests itself for large time intervals which are not covered by the simple TLS model presented in Fig. 3. It should be noted that the analysis of trails of several dozens of individual Terylene molecules in polyethylene, carried out in [14], indicates that, indeed, some of the encountered cases cannot be described by the TLS model (e.g., when the optical line of a molecule jumps between three spectral positions). The factors responsible for the auxiliary broadening in Fig. 7b remain unclear.

7. CONCLUSIONS

The above analysis leads to the following explanation to the peculiar time dependence of the half-width of the lines of three individual Terylene molecules in polyethylene (see Fig. 1).

The weak logarithmic increase in the half-width, which is the same for the three lines, is explained by the interaction of chromophore molecules with a vast number of TLS's of the polymer (including many nonequilibrium systems). This effect is usually referred to as spectral diffusion. It was observed more than once earlier in the study of spectral holes and optical lines corresponding to individual molecules.

The sharp increase in the line half-width of Terylene molecule *2* in Fig. 1 is explained satisfactorily by an effect similar to the effect of spectral line saturation observed for molecules with a large quantum yield in the triplet state of a molecule. It is only in this case that the role of a triplet trap is played by the second state of the TLS simulating the two possible conformation states of an impurity center. The transition to the second conformation state of the impurity center upon the absorption of a photon is precisely responsible for the sharp line broadening. If the mean length of *on* and *off* intervals appearing in formulas (34) and (35) were measured in [14], these formulas would not contain free parameters. Unfortunately, the mean lengths of *on* and *off* intervals have not been measured in [14]. For this reason, these parameter had to be selected. The processing of the measured time dependences of the half-widths of the lines emitted by molecule *2* made it possible to determine the tunneling rates $R = 1/\tau_{off}$ and B in the ground and excited electron state of molecule *2*, respectively.

Figure 7 shows that the simultaneous sharp change in the population and half-width of the line of molecule *3* is also caused by light-induced tunneling. However, it cannot be described completely by the simple TLS model since this molecule exhibits an auxiliary broadening of unclear nature for large values of time.

ACKNOWLEDGMENTS

The authors is grateful to L. B. Ershova and E. V. Khots for their help in preparing figures.

This work was supported financially by the Russian Foundation for Basic Research, project no. 01-02-16580.

REFERENCES

1. I. S. Osad'ko, in *Spectroscopy and Excitation Dynamics of Condensed Molecular Systems*, Ed. by V. Agranovich and R. Hochstrasser (North-Holland, Amsterdam, 1983; Nauka, Moscow, 1987).
2. K. Duppen, L. W. Molenkamp, J. B. L. Morsink, *et al.*, Chem. Phys. Lett. **84**, 421 (1981).
3. W. Breinl, J. Friedrich, and D. Haarer, J. Chem. Phys. **81**, 3915 (1984).
4. K. Littau, Y. S. Bay, and M. D. Fayer, J. Chem. Phys. **92**, 4145 (1990).
5. H. C. Meijers and D. A. Wiersma, J. Chem. Phys. **101**, 6927 (1994).
6. S. Zilker and D. Haarer, Chem. Phys. **220**, 167 (1997).
7. W. B. Mims, K. Nassau, and J. D. McGee, Phys. Rev. **123**, 2059 (1961).
8. H. Maier, B. Kharlamov, and D. Haarer, in *Tunneling Systems in Amorphous Solids*, Ed. by P. Esquinazi (Springer-Verlag, Berlin, 1998).
9. L. Fleury, A. Zumbusch, M. Orrit, *et al.*, J. Lumin. **56**, 15 (1993).
10. R. Kettner, J. Tittel, Th. Basche, and C. Brauchle, J. Phys. Chem. **98**, 6671 (1994).
11. R. Klauder and P. W. Anderson, Phys. Rev. **125**, 912 (1962).
12. I. S. Osad'ko, Zh. Éksp. Teor. Fiz. **109**, 805 (1996) [JETP **82**, 434 (1996)].
13. I. S. Osad'ko, Zh. Éksp. Teor. Fiz. **116**, 962 (1999) [JETP **89**, 513 (1999)].
14. A. M. Boiron, Ph. Tamarat, B. Lounis, *et al.*, Chem. Phys. **247**, 119 (1999).
15. T. L. Reineke, Solid State Commun. **32**, 1103 (1979).
16. E. Geva and J. L. Skinner, J. Chem. Phys. **109**, 4920 (1998).
17. I. S. Osad'ko and N. N. Zaitsev, Chem. Phys. Lett. **293**, 138 (1998).
18. I. S. Osad'ko and L. B. Yershova, J. Chem. Phys. **112**, 9645 (2000).
19. I. S. Osad'ko, *Selective Spectroscopy of Single Molecules* (Fizmatlit, Moscow, 2000).
20. I. S. Osad'ko, Zh. Éksp. Teor. Fiz. **113**, 1606 (1998) [JETP **86**, 875 (1998)].
21. J. Bernard, L. Fleury, H. Talon, and M. Orrit, J. Chem. Phys. **98**, 850 (1993).
22. M. Nirmal, B. O. Dabbousi, M. G. Bawendi, *et al.*, Nature **383**, 802 (1996).
23. M. E. Pistol, P. Castrilo, D. Hessman, *et al.*, Phys. Rev. B **59**, 10725 (1999).
24. H. P. Lu, L. Xun, and X. S. Xie, Science **282**, 1877 (1998).
25. X. S. Xie and H. P. Lu, J. Biol. Chem. **274**, 15962 (1999).
26. O. N. Korotaev, E. I. Donskoĭ, V. I. Glyadkovskiĭ, and V. N. Kopranenkov, Opt. Spektrosk. **57**, 145 (1984) [Opt. Spectrosc. **57**, 88 (1984)].
27. W. E. Moerner, T. Plakhotnik, T. Irgartinger, *et al.*, J. Phys. Chem. **98**, 7382 (1994).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 93, No. 4, 2001, pp. 876–894.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 1004–1026.

MISCELLANEOUS

Relativistic Quantum Protocols: Bit Commitment and Coin Tossing

S. N. Molotkov* and S. S. Nazin

Institute of Solid State Physics, Russian Academy of Sciences, p/o Chernogolovka, Moscow oblast, 142432 Russia
**e-mail: molotkov@issp.ac.ru*

Received November 21, 2000

Abstract—New relativistic quantum protocols realizing the bit commitment and coin tossing schemes are proposed. The protocols are based on the idea that spatially extended nonstationary orthogonal quantum states inaccessible for measurement cannot be unambiguously distinguished. As the states are transmitted from the region controlled by one party to the region accessible for measurement by the other party, the states become reliably distinguishable when accessed as a whole. Essential points of the protocol are both the quantum character of states and the existence of an ultimate signal propagation speed dictated by special relativity. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In the nonrelativistic quantum mechanics, any measurement performed by an observer in a quantum system introduces, in the general case, a disturbance in the initial state of the system. In the absence of limitations with respect to the maximum speed, instantaneous measurements (including spatially nonlocal) are formally not prohibited.

The special relativity theory poses additional restrictions on the measurements in the classical sense. The existence of a limiting speed makes the nonlocal measurements on a classical object within an arbitrarily short time impossible.

The relativistic quantum field theory must involve additional restrictions concerning the measurement of observables in comparison with the nonrelativistic quantum mechanics. Apparently, the problem of fundamental limitations imposed by special relativity upon the possibility of measuring various dynamic variables in a quantum system was originally considered by Landau and Peierls [1]. The qualitative considerations formulated in that paper based on the uncertainty relationships, together with the restriction of ultimate speed led to a conclusion that "…all physical quantities employed by the wave mechanics turn out to be indeterminable in the relativistic field."

The nonrelativistic quantum mechanics does not formally prohibit precisely measuring, for example, the momentum of a quantum system within an arbitrarily short (zero) period of time. The momentum operator is nonlocal and the eigenstate of this operator (i.e., a plane wave) is infinitely extended in space. Of course, being not normalized, the plane wave is not a physically realizable state. However, this state can be approached with any precision by a state (localized in an arbitrarily large but still finite region of space) such that the average value of the momentum operator measured in this state can be arbitrarily close to the momentum of the plane wave. This momentum measurement implies that the state occupying the arbitrarily large region in space can be accessed as a whole. The nonrelativistic quantum mechanics does not prohibit the access to any spatial region within an arbitrarily short (zero) period of time and, hence, does not poses formal restrictions on the arbitrarily accurate instantaneous measurement of, for example, the momentum. With an allowance for limitations of special relativity, the access to an infinite spatial region requires an infinite time: in this sense, the physical quantities are indeterminate (more strictly speaking—indeterminable, if the exact determination within a finite time is implied).

From the standpoint of precise measurement of the momentum (using an eigenstate of the momentum operator or, more precisely, generalized eigenvector), this quantity is indeterminable even in the nonrelativistic case because the plane wave state cannot be realized exactly. We may only approach, with an arbitrary accuracy, the precise measurement. It is essential that the nonrelativistic quantum mechanics poses no restrictions on the time required to obtain the result. The relativistic theory also does not prohibit the measurement of, for example, the momentum with arbitrarily high precision. However, the need in accessing the formally infinite region for such a precise measurement of the momentum implies an infinite period of time. Interpreted in this way, the statement of Landau and Peierls [1] concerning the impossibility of precisely determining the momentum is self-consistent.

The problem of measurements in quantum systems in the relativistic case was further investigated by Bohr and Rosenfeld [2]. As far as we can judge, the conclusions derived in [2] do not abolish the restrictions indi-

cated in [1] concerning the necessary time because these considerations follow essentially from limitations dictated by the special relativity. Later, the arguments formulated in [1] were reproduced virtually unchanged in [3].

Below we will be interested in restrictions posed by the relativistic quantum field theory on the probability of unambiguously distinguishing between two states of a quantum system. Since the information is carried by photons (representing essentially relativistic particles), the relativistic limitations of measurements are of principal significance. In addition, the restrictions with respect to the ability of distinguishing quantum states may only increase the possibility of constructing various relativistic cryptography protocols.

All the nonrelativistic quantum cryptography protocols employ, in one or another form, the following two circumstances. The first is the so-called no-cloning theorem [4], according to which an unknown quantum state cannot be duplicated. This is equivalent to impossibility of the process

$$|A\rangle|\psi\rangle \longrightarrow U(|A\rangle|\psi\rangle) = |B_\psi\rangle|\psi\rangle|\psi\rangle,$$

where $|A\rangle$ and $|B_\psi\rangle$ are the states of the system before and after duplication and U is some unitary operator. This process is prohibited by virtue of the linearity and unitarity of evolution in the quantum mechanics. The second point is that even a weaker process, delivering any information about one of the two nonorthogonal states without disturbing these states, is impossible [5]. This circumstance prohibits processes of the type

$$|A\rangle|\psi_1\rangle \longrightarrow U(|A\rangle|\psi_1\rangle) = |A_{\psi_1}\rangle|\psi_1\rangle,$$

$$|A\rangle|\psi_2\rangle \longrightarrow U(|A\rangle|\psi_2\rangle) = |A_{\psi_2}\rangle|\psi_2\rangle,$$

with $|A_{\psi_1}\rangle \neq |A_{\psi_2}\rangle$, provided that $\langle\psi_1|\psi_2\rangle \neq 0$, which implies the impossibility of distinguishing nonorthogonal states. No such a prohibition exists with respect to orthogonal states. Moreover, the nonrelativistic quantum mechanics, generally speaking, does not prohibit instantaneous (arbitrarily fast) unambiguous distinguishing between the orthogonal states at any time instant without disturbing these states. If there is a pair of orthogonal states in the Hilbert space $\mathcal{H}$ ($|\psi_{1,2}\rangle \in \mathcal{H}$, $\langle\psi_1|\psi_2\rangle = 0$), these states can be reliably distinguished using the measurement described by an orthogonal expansion of unity in the space $\mathcal{H}$:

$$\mathcal{P}_1 + \mathcal{P}_2 + \mathcal{P}_\perp = I, \quad \mathcal{P}_{1,2} = |\psi_{1,2}\rangle\langle\psi_{1,2}|,$$

$$\mathcal{P}_\perp = I - \mathcal{P}_1 - \mathcal{P}_2,$$

where $\mathcal{P}_{1,2}$ are the projectors onto subspaces $\mathcal{H}_{1,2}$ generated by $|\psi_{1,2}\rangle$; $\mathcal{P}_\perp$ is the projector onto subspace $\mathcal{H}_{1,2}^\perp = (\mathcal{H}_1 \oplus \mathcal{H}_2)^\perp$. The probability of obtaining a result belonging to the manifold of results $\Theta = \{1, 2, \perp\}$ (e.g., for the initial state $|\psi_1\rangle$) is

$$\Pr\{|\psi_1\rangle\} = \mathrm{Tr}\{|\psi_1\rangle\langle\psi_1|\mathcal{P}_1\} = 1, \tag{1}$$

while that in the $\mathcal{P}_{2,\perp}$ channel is identically zero

$$\Pr\{|\psi_1\rangle\} = \mathrm{Tr}\{|\psi_1\rangle\langle\psi_1|\mathcal{P}_{2,\perp}\} = 0, \tag{2}$$

(analogous expressions can be written for the initial state $|\psi_2\rangle$). These relationships imply that the orthogonal states are unambiguously distinguishable. Moreover, these states can be distinguished instantaneously and without disturbing the system. It is assumed that the whole Hilbert space is accessible at once. Since the physical systems cannot exist outside the coordinate space, the accessibility of the Hilbert space of states implies the access to the region of the coordinate space in which the state carrier is nonzero.

In the relativistic case, the impossibility of instantaneously accessing a finite space region results in that the orthogonal states extended in the coordinate space can be unambiguously distinguished as proper objects only when accessible as a whole (i.e., the measuring instrument can access the entire spatial region in which the state carriers are nonzero). This circumstance will be employed below in constructing the relativistic quantum protocol.

Many problems in cryptography can be reduced to a sequence of primitive cryptographic exchange protocols. These are the protocol of key transmission (cryptography) [6–8], bit commitment, and coin tossing [9–12]. The bit commitment represents a stronger protocol as compared to the coin tossing in a sense that, once the bit commitment is realized, the coin tossing protocol can be formulated on this basis.

In a descriptive and informal way, the bit commitment protocol is usually formulated as follows. There are two spatially separated legal parties of the protocol, A and B (henceforth conventionally referred to as "Alice" and "Bob"). In the commitment stage, Alice selects a secret bit (0 or 1) and transfers a part of information about this to Bob; by this part, Bob cannot unambiguously judge on what Alice means. More strictly speaking, the probability of identifying the message by measuring the part of information in the ideal case is 1/2 (the probability of simple guessing). In the stage of unveiling, Bob may request the remainder bit information from Alice. In the ideal case, Bob must unambiguously (with the probability of unity) recover in the unveiling stage the secret bit originally committed by Alice. Moreover, Alice must be unable of changing the secret bit when the first part of information is sent to Bob.

The protocol of coin tossing is formulated as follows. Two spatially separated parties (neither of the two trusting another, each party having all admissible means of deceiving another) must exchange information so that eventually (in the ideal case, at the probability of unity) they both have to agree that the bit trans-

mitted by this protocol is a true choice. If the parties could exchange information only via a classical communication channel, this task might even seem unsolvable.

It is evident that, once the secret bit commitment protocol can be realized, the coin tossing protocol can be implemented on this basis as well. To this end, Bob has to guess the secret bit (that was originally committed by Alice and will be unambiguously known to both upon unveiling) during the time interval between the first bit commitment and unveiling. If Bob's guess is correct, he wins, otherwise Alice wins.

A simple variant of the bit commitment protocol is offered by the following example. Alice writes the secret bit on a sheet of paper, places it into a safe box, and sends the box to Bob (commitment), while retaining the key till the unveiling stage. Despite evident simplicity, this example reflects all principal features of the protocols employing the classical information carriers. Here, Bob receives at his disposal essentially the whole information (rather than a part of it) concerning the secret bit. The laws of nature do not forbid Bob (provided that the appropriate technical facilities are available) to know the secret data before official unveiling. An analogous situation takes place in the protocols based on the computational complexity of some trap functions such as the discrete logarithm [13]. In these protocols, Alice transmits to Bob the y value ($y = a^b \bmod p$; a and p being known preliminarily and party b being the secret bit). The information about b (provided with the y value) is already sufficient to unambiguously establish (by calculating the discrete logarithm) the secret bit value. However, from the standpoint of the classical computational algorithms, such calculations would require large (exponentially increasing) computational facilities. However, it is not proved by now that other, more effective (polynomial) classical algorithms do not exist.

In the case when Alice and Bob may exchange information only via a classical communication channel, the problem was solved by Blum [13]. Strictly speaking, the protocol proposed in [13] is not secure with respect to the possibility of one party being deceived by another, since the secrecy is based on the unproved difficulty of computing the discrete logarithm. In protocols of this type, Alice also transmits the whole information (rather than a "part" of it), and the laws of nature do not forbid Bob to know the secret bit before the unveiling stage (e.g., by using a quantum computer [14, 15], although this device is still far from practical realization).

Using classical (nonrelativistic) objects as information carriers, it is impossible to realize an unconditionally secure bit commitment protocol (the secrecy of which is based only on the fundamental natural prohibitions, rather than on restricted technical facilities), whereby only a "part" of the classical object (e.g., of a spatially extended signal) is delivered to Bob prior to the unveiling sage. Since the signal part accessible to Bob prior to unveiling must appear identically for the bit values 0 and 1 (so that Bob would possess zero information about bit commitment before unveiling), the object (signal) part retained by Alice must differ for the two values. The laws of nonrelativistic classical physics do not prohibit instantaneous modification of the retained signal part (to change from 0 to 1 and vice versa) before sending it to Bob for unveiling. Thus, nothing forbids Alice to change the secret bit value. Thus, an unconditionally secure bit commitment protocol cannot be realized within the framework of the nonrelativistic classical physics.

The nonrelativistic quantum protocols employ quantum systems as the information carriers. In general, the common features of these protocols are as follows. Let $\mathcal{H}_s$ be the Hilbert space containing the states of information carriers. Alice selects the system state $|\psi_{0,1}\rangle \in \mathcal{H}_s$ corresponding to 0 or 1 and sends these states to Bob (as a rule, the nonorthogonal states are selected). It is important to note that the whole space of states $\mathcal{H}_s$ is implicitly assumed to be accessible to both Alice and Bob during the entire protocol time. The requirement that the matrices of density of states corresponding to 0 and 1 would appear the same for Bob prior to unveiling opens the possibility for Alice to deceive Bob by the undisclosed Einstein–Podolsky–Rosen (EPR) attack [16, 17]. Roughly speaking, this protocol involves only the space of the system states. However, this situation is rather artificial and does not correspond to real conditions of the information transfer. More precisely, we imply that the protocol parties cannot monitor the entire space and only control certain regions (laboratories, measuring devices, etc.). Moreover, all measurements take pace in the real space and time (or space–time in the relativistic case). The nonrelativistic quantum mechanics does not prohibit the formation of intricate states in physically distinguishable systems (we are interested only in such cases since, should the systems be identical, it is impossible to perform a measurement involving only one of these systems). Therefore, if the protocol parties control only nonintersecting regions of the space, an intricate state from $\mathcal{H}_s \oplus \mathcal{H}_a$ must automatically be nonlocal in the coordinate space. The wave functions of both subsystems from $\mathcal{H}_s$ and $\mathcal{H}_a$ must possess the carriers simultaneously in the regions controlled by both Alice and Bob (otherwise the given state will not be intricate for the parties). The latter condition implies that each party has access both to the space of states $\mathcal{H}_s$ and to $\mathcal{H}_a$, being capable of conducting measurements and performing unitary transformations in the subsystems (by virtue of their being physically distinguishable) at its own discretion. Thus, the local character of transformations in the space of states $\mathcal{H}_s \oplus \mathcal{H}_a$ (involving manipulations in one of the subspaces of states) does not imply locality in the coordinate space. In other words, the space of states of the information carriers in

such nonrelativistic quantum protocols (declining from explicitly taking into account the space–time structure of the information carriers) is accessible to both parties. In this sense, the protocols do not employ the idea of delivering a part of information about the secret bit carrier.

Apparently, an explicit use of the effects of state transmission in the coordinate space, when Bob can access only a part of the state (because of its extension in space) in the nonrelativistic case, cannot bring some new essential features of the problems under consideration, since there is no limiting speed of the signal propagation.

It should be noted that the problem formulation using only specific features of the space of states $\mathcal{H}$ does not correspond to a real situation taking place in the case of information transmission in the real space–time. It is more natural to consider a problem in which both protocol parties reside in their laboratories and control the space regions in the vicinity of these laboratories. Obviously, Alice and Bob can neither control nor monitor the whole coordinate space simultaneously.

For the first time, an explicit use of the effects of the state propagation (with the space–time structure of states taken into account) in solving the tasks of quantum cryptography was suggested by Goldenberg and Vaidman [18] (in our opinion, the ideas formulated in that paper were not timely appreciated [19, 20]). An allowance for the restrictions posed by the special relativity and quantum mechanics (quantum field theory) considerably simplifies the proof of unconditional security of the relativistic quantum cryptosystems [21, 22]. Moreover, the quantum field theory poses additional fundamental limitations, for example, on the teleportation of quantum states [23].

Recently, Kent [24] suggested the classical protocols of bit commitment and coin tossing with an allowance for the limited velocity of signal (information) propagation. The relativistic classical protocol [24] is unconditionally secure (i.e., the protocol security is based only on the basic laws of nature) and allows, in principle, the second stage of the protocol to be delayed (i.e., to retain information about the secret bit committed by Alice for an arbitrarily long time). For the protocol realization, it is necessary that both parties control two spatially-separated sites.

An idea of the relativistic quantum bit commitment and coin tossing protocols using orthogonal states was previously formulated in [25] based on two simple considerations. First, a pair of orthogonal states (unambiguously distinguishable if accessible for measurement as a whole) become effectively nonorthogonal (undistinguishable) when bounded in a subspace. This circumstance is valid in the nonrelativistic quantum mechanics as well. Indeed, a pair of orthogonal states extended in the coordinate space, for example,

$$(\psi_0, \psi_1) = \int_{-\infty}^{\infty} \psi_0^*(x)\psi_1(x)dx = 0,$$

$$\psi_{0,1}(x) \in \mathcal{L}^2(-\infty, \infty, dx),$$

becomes effectively nonorthogonal when bounded in a subspace (a finite region Ω in the coordinate space):

$$(\psi_0, \psi_1)_\Omega = \int_\Omega \psi_0^*(x)\psi_1(x)dx \neq 0.$$

The second important point is the presence of a limited velocity of propagation of for quantum states and classical objects, as dictated by the special relativity. This condition ensures that the access to the whole state (in the region where it exists) cannot be instantaneous.

Using the states with "internal" degrees of freedom (e.g., helicity of the photon) allows the protocols to be simplified in comparison with those considered in [25], where the states possessed no such degrees of freedom. Strictly speaking, the states for 0 and 1 of the type

$$|\psi_{0,1}\rangle = \psi(x) \otimes |e_{0,1}\rangle, \quad \langle e_0|e_1\rangle = 0,$$

are orthogonal (due to the internal degrees of freedom) even if not accessible for the measurement as a whole (but only in that part of the coordinate space where the function $\psi(x)$ is nonzero). However, these states cannot be unambiguously distinguished: the probability of this (i.e., of measuring 0 or 1 in one of the orthogonal channels) can be arbitrarily small if only a part of the state accessible. Indeed, the probability of measurement in a finite region of the space Ω

$$\langle \psi_i|\psi_i\rangle_\Omega = \int_\Omega |\psi(x)|^2 dx < 1, \quad i = 0, 1,$$

can be rendered arbitrarily small by selecting the shape and size of $\psi(x)$. This is in fact dictated by the condition of normalization of the state

$$\int_{-\infty}^{\infty} |\psi(x)|^2 dx = 1.$$

Our idea of the relativistic quantum protocol is generally as follows. Alice, controlling a certain finite region of the coordinate space, prepares a quantum state at a time instant stipulated by the protocol. This state propagates in the communication channel and gradually becomes accessible for Bob's measurement in a region not controlled by Alice. Being in access only to a part of the quantum state in the real space, Bob cannot obtain reliable information about the secret bit (i.e., unambiguously distinguish 0 and 1); the states can be selected so that the probability of distinguishing 0 and 1 by measurement on the Bob side would be arbitrarily

close to 1/2 (the probability of simple guessing) within an arbitrarily long time interval (initially stipulated by the protocol). There are no fundamental limitations from above for this interval (although making this interval arbitrarily long can be technically difficult). The existence of limiting velocity allows the states to be selected so that, when a part of the state leaves the region controlled by Alice and becomes out-of-reach, she cannot change this state (committed bit). As the protocol time elapses, the whole states become accessible to Bob and he can obtain unambiguous information at a probability arbitrarily close to unity. This protocol is based essentially on both the quantum character of states and the existence of a limiting velocity, as dictated by the special relativity.

Section 2 describes the states and measurements involved in the proposed protocols. Sections 3 and 4 present the bit commitment and coin tossing protocols for the states with compact carriers. Section 5 describes changes in the protocols, which are necessary for taking into account the principal nonlocality of states in the quantum field theory. The Conclusion briefly summarizes the results of this study.

2. STATES AND MEASUREMENT INVOLVED IN THE PROTOCOL

Since the protocol explicitly employs the space–time structure of states, the procedure cannot be formulated without taking into account a particular geometry of the system. Let us consider a one-dimensional model that retains all essential features of the protocol dictated by the quantum field theory. This model is frequently employed in the quantum optics. We will deal with the field states corresponding to a massless field, the states of which a determined in the momentum representation on the mass surface $k_0^2 - k^2 = 0$. We are interested in the states propagating in the positive direction of the x axis ($k > 0$). It is assumed that Alice controls a region in the vicinity of the point x_A, while Bob controls an analogous region in the vicinity of x_B ($x_A < x_B$).

All the functions considered below depend on the difference $\tau = t - x$; the speed of light is taken equal to unity ($c = 1$). This representation reflects intuitive notions of a wave packet moving at the speed of light. The eigenvector $|k\rangle$ of the momentum operator, corresponding to the momentum k, is a generalized eigenvector [more precisely, a linear continuous functional on elements of a dense subset in $\mathcal{H} = \mathcal{L}^2(0, \infty, d\xi)$] having the following form:

$$\langle \xi | k \rangle = \delta(k - \xi). \tag{3}$$

The states $|\psi\rangle$ belonging to $\mathcal{H}$ can be expanded in generalized states

$$|\psi\rangle = \int_0^\infty \langle k|\psi\rangle |k\rangle dk, \tag{4}$$

where the values of the functional $\langle k|$ on the elements $|\psi\rangle$ are

$$\langle k|\psi\rangle = \int_0^\infty \psi(\xi)\delta(k - \xi)d\xi = \psi(k),$$

(representing the amplitude of the state $|\psi\rangle$ in the k- representation). Accordingly, the amplitude of the state $\langle k|$ in the τ-representation has the following form:

$$\langle k|\tau\rangle = \frac{1}{\sqrt{2\pi}} e^{ik\tau}, \quad k \in (0, \infty), \quad t \in (-\infty, \infty), \quad \tau = t - x. \tag{5}$$

This form reflects intuitive notions of a plane wave (representing the state with a certain momentum) moving at the speed of light.

An essential point in the protocol is that the quantum states are propagating at the maximum possible speed (light speed). In the τ-representation, the orthogonal states (packets) corresponding to 0 and 1 in the protocol have the following form:

$$|\psi_{0,1}\rangle = \int_{-\infty}^\infty f(\tau)|\tau\rangle d\tau \otimes |e_{0,1}\rangle, \quad \langle e_0|e_1\rangle = 0, \quad \langle e_{0,1}|e_{0,1}\rangle = 1, \tag{6}$$

where the states $|e_{0,1}\rangle$ correspond to the internal degrees of freedom (e.g., helicity of the photon).

The condition of normalization for the states is

$$\langle \psi_{0,1}|\psi_{0,1}\rangle = \int_{-\infty}^\infty \int_{-\infty}^\infty f(\tau) f^*(\tau')\langle \tau|\tau'\rangle d\tau d\tau' = 1, \tag{7}$$

where

$$\langle \tau|\tau'\rangle = \frac{1}{2\pi}\delta_+(\tau - \tau') = \frac{1}{2\pi}\int_0^\infty e^{ik(\tau - \tau')}dk = \frac{1}{2}\delta(\tau - \tau') + \frac{i}{\pi}\frac{1}{\tau - \tau'}. \tag{8}$$

Let us introduce the state amplitude in the k-representation, which can be defined as

$$f(\tau) = \int_0^\infty f(k)e^{-ik\tau}dk. \tag{9}$$

Taking into account expressions (5)–(7), the normalization condition can be written as

$$\langle \psi_{0,1}|\psi_{0,1}\rangle = \int_{-\infty}^\infty \int_{-\infty}^\infty f(\tau) f^*(\tau') \times \left[\frac{1}{2}\delta(\tau - \tau') + \frac{i}{\pi}\frac{1}{\tau - \tau'}\right] d\tau d\tau'. \tag{10}$$

Substituting the amplitude in the k-representation from (7) into (8) and taking into account that [26]

$$\int_{-\infty}^{\infty} e^{ik\tau}\frac{1}{\tau + a}d\tau = i\pi\,\mathrm{sgn}\,k e^{-iak}, \tag{11}$$

we obtain

$$\langle\psi_{0,1}|\psi_{0,1}\rangle = \int_{-\infty}^{\infty}|f(\tau)|^2 d\tau = 1. \tag{12}$$

According to the requirement of microscopic causality [27], the field operators (generating the field states belonging to the Hilbert space of states by acting upon the vacuum vector) must be commutative (or anticommutative) if these operators belong to spacelike regions. As is known, the commutator of the field operators is a generalized function (for detail, see [27]). In order to speak of the local properties of the generalized functions, the basic functions must possess certain properties (in fact, belonging to the space $\mathcal{S}(\hat{x})$ of infinitely smooth functions decreasing at the infinity more rapidly than $|\hat{x}|^{-n}$ for any natural n). In other words, the states of the free field cannot possess a compact carrier (i.e., differ from zero only in a finite spatial region and belong to the space $\mathcal{D}(\hat{x})$), that is, the free field states are fundamentally nonlocalizable. However, the field theory allows the existence of states arbitrarily localized in space and decreasing at a rate arbitrarily close to exponential (see, e.g., [28–32]). In addition, the functions from space $\mathcal{D}(\hat{x})$ with a compact carrier form a dense set in the space $\mathcal{S}(\hat{x})$, which implies that any function from $\mathcal{S}(\hat{x})$ can be approximated by the functions from $\mathcal{D}(\hat{x})$ with an arbitrary accuracy.

In application to the one-dimensional model under consideration, the property of nonlocalizability can follow from the Wiener–Paley theorem [33], since the normalization condition (10) with an allowance for (7) implies a quadratic integrability of the amplitude in the k-representation and poses restrictions on the asymptotic behavior of the function $f(\tau)$

$$f(\tau) = \int_0^{\infty} f(k)e^{-ik\tau}dk, \quad \int_{-\infty}^{\infty}\frac{|\ln|f(\tau)||}{1+\tau^2}d\tau < \infty. \tag{13}$$

As follows from (11), the function $f(\tau)$ cannot possess a carrier that is compact with respect to τ and cannot decrease exponentially, but it can be made arbitrarily strongly localized and decreasing at a rate arbitrarily close to exponential. For example:

$$f(\tau) \propto \exp\left\{-\frac{\alpha\tau}{\ln(\ln(\ldots\ln\tau))}\right\}, \tag{14}$$

where α is an arbitrary positive quantity.

For the sake of better illustration and convenience, we will first formulate a protocol for the states with compact carrier (since the functions from $\mathcal{D}(\tau)$ form a dense set and any function $f(\tau)$ can be approximated with an arbitrary accuracy by the functions from $\mathcal{D}(\tau)$). Then we will introduce changes in the protocols, which are necessary for nonlocalizable states.

Let the state carrier $f(\tau)$ possess a compact carrier $\mathrm{supp}\,f(\tau) = (-\Delta\tau, \Delta\tau)$ (where $\Delta\tau$ can be selected arbitrarily small). The states are formed with contributions only from the vectors $|\tau\rangle$ from the region $(-\Delta\tau, \Delta\tau)$ on the light cone

$$|\psi_{0,1}\rangle = \int_{-\Delta\tau}^{\Delta\tau} f(\tau)|\tau\rangle d\tau \otimes |e_{0,1}\rangle. \tag{15}$$

The situation with the quantum field theory differs from the nonrelativistic quantum protocols, in which the space–time structure of states is not explicitly employed and the effects of state preparation are not significant [more strictly speaking, the nonrelativistic quantum mechanics does not prohibit instantaneous (at any time moment) preparation of states from $\mathcal{H}$, even including the states that are nonlocal in the coordinate space]. In contrast to this, the preparation of states in the field theory requires the access to a finite region of the space–time (even if the carrier is compact). In the one-dimensional case, this corresponds either to a coordinate space region with the size $\Delta x = 2\Delta\tau$ (if the state is prepared using a nonlocal source at a fixed time moment t) or to a finite time interval $\Delta t = 2\Delta\tau$ (if the state is generated by a point source located at the point x). Therefore, a relativistic protocol can only be formulated with an allowance for a particular geometry of the system. The one-dimensional case is most simple for the analysis, since all quantities depend on a single variable $\tau = x - t$. Taking into account that there are experiments employing quasi-one-dimensional fiber-optic systems, an analysis of the proposed one-dimensional model is quite reasonable.

Now we will consider the "extended" states (required for the protocol) composed of two halves spaced by an interval τ_0 on the light cone:

$$|\psi_{0,1}(\tau_0)\rangle = \frac{1}{\sqrt{2}}\int(f(\tau) + f(\tau-\tau_0))|\tau\rangle d\tau \otimes |e_{0,1}\rangle. \tag{16}$$

Here and below, the states are normalized as

$$\int_{-\infty}^{\infty}|f(\tau)|^2 d\tau = \int_{-\infty}^{\infty}|f(\tau-\tau_0)|^2 d\tau = 1, \tag{17}$$
$$\mathrm{supp}\,f(\tau)\cap\mathrm{supp}\,f(\tau-\tau_0) = \backslash\,\mathrm{varnothing}.$$

Since the initial state possesses a carrier in $(-\Delta\tau, \Delta\tau)$, the state preparation requires controlling the region $(-\Delta\tau, \Delta\tau + \tau_0)$ on the light cone [representing either the

region $\Delta x = (-\Delta\tau, \Delta\tau + \tau_0)$ in the coordinate space if the state is prepared by a nonlocal instrument at a fixed time moment, or a time interval $\Delta t = (-\Delta\tau, \Delta\tau + \tau_0)$ if the state is generated by a local source at the point x].

Now let us discuss the individual measurements performed by the second party of the protocol on a separate state. The measurements are described by an expansion of unity (with the results of these measurements forming the space $\Omega = \{\tau \in (-\infty, \infty), i = 0, 1\}$), for which

$$I = \left(\int_{-\infty}^{\infty} \mathcal{M}(d\tau)\right) \otimes (\mathcal{P}_0 + \mathcal{P}_1)$$
$$= \left(\int_{-\infty}^{\infty} d\tau \left(\int_0^{\infty} e^{ik\tau}|k\rangle dk\right)\left(\int_0^{\infty} e^{-ik'\tau}\langle k'|dk'\right)\right) \otimes (\mathcal{P}_0 + \mathcal{P}_1), \quad (18)$$
$$\langle k|k'\rangle = \delta(k - k'),$$

where $|k\rangle$ is the formal eigenvector corresponding to a preset k and

$$\mathcal{M}(d\tau) = |\tau\rangle\langle\tau|d\tau, \quad \mathcal{P}_0 = |e_0\rangle\langle e_0|, \quad \mathcal{P}_1 = |e_1\rangle\langle e_1|.$$

It will be also necessary to describe the propagation of states in a quantum communication channel from the region controlled by Alice to that controlled by Bob. This propagation is described by a unitary translation of the state $|\psi_{0,1}\rangle$ along the branch of the light cone $\tau = x - t$:

$$\mathcal{U}_{\text{ch}}(\tau_{\text{ch}})|\psi_{0,1}\rangle = |\psi_{0,1,\tau_{\text{ch}}}\rangle = \frac{1}{\sqrt{2}}$$
$$\times \int (f(\tau - \tau_{\text{ch}}) + f(\tau - \tau_0 - \tau_{\text{ch}}))d\tau \otimes |e_{0,1}\rangle, \quad (19)$$

where τ_{ch} is the length of the communication channel. A relationship between the state "extension" $(2\Delta\tau + \tau_0)$ and the channel length τ_{ch} must be such that

$$\tau_{ch} < \tau_0 + 2\Delta\tau,$$

which allows us to assume without loss of generality that $\tau_{\text{ch}} = 0$ (since the channel length can be arbitrarily short, only not exceeding the state "extension").

The probability for Bob to obtain a result in the ith channel in the $d\tau$ interval for an input state $|\psi_j(\tau_0)\rangle$ is

$$\Pr\{d\tau; i, j\}$$
$$= \text{Tr}\{(\mathcal{M}(d\tau) \otimes \mathcal{P}_i)|\psi_j(\tau_0)\rangle\langle\psi_j(\tau_0)|\} \quad (20)$$
$$= \delta_{ij}\frac{1}{2}\{|f(\tau)|^2 + |f(\tau - \tau_0)|^2\}d\tau.$$

This expression describes the density of the probability of obtaining the result in one of the orthogonal (distinguishable) channels for 0 ($i = j = 0$) and 1 ($i = j = 1$) in the interval $d\tau$. On the intuitive level, this measurement can be interpreted as realizable using a photodetector operating in a standby mode with a small (formally zero) internal response time. The result of this measurement is a random event falling within the interval $d\tau$ at a probability density (20).

The probability of detecting a state within a finite interval $\Delta(\tau)$ (for $i = j$) is

$$\Pr\{\Delta(\tau)\} = \int_{\Delta(\tau)} \Pr\{d\tau; i, i\}$$
$$= \frac{1}{2}\left\{\int_{\Delta(\tau)} |f(\tau)|^2 d\tau + \int_{\Delta(\tau)} |f(\tau - \tau_0)|^2 d\tau\right\}. \quad (21)$$

If the interval $\Delta(\tau)$, representing accessible region on the light cone, does cover the whole state carrier (e.g., covering one half of this state), the probability of obtaining the result is 1/2. However, once the result is obtained, the orthogonality of the registration channels $\mathcal{P}_0$ and $\mathcal{P}_1$ provides for the unambiguous identification of states. For this reason, the probability of an error in the identification of state by measurement during the time $\Delta\tau \le \tau \le \tau_0 + \Delta\tau$ is 1/4. Accordingly, the probability of the correct identification is 3/4. For a simple guessing without measurement, the error is 1/2.

It should be emphasized that this measurement can by no means be interpreted as lasting for a finite time $\Delta(\tau)$. Every time, the result of the measurement appears randomly at a time moment t with the probability density (20). On elapse of the time $\tau_0 + 2\Delta\tau$, when the state occurs completely in the region accessible for Bob's measurement, 0 or 1 are unambiguously identified due to the orthogonality of states.

Thus, the propagation of states with ultimately possible speed allows the idea of a partial delivery of the information concerning the committed secret bit (a part of the quantum state) by Alice to be explicitly and naturally realized. The quantum character of states is also essential for the protocol. Indeed, for $f(\tau)$ describing the shape of a signal with different polarizations (e_0 or e_1), the probability of the identification of states in the classical case (even provided the access to a half of the state) would be 1 (against 3/4 in the quantum case). In the quantum case, the probability of 3/4 obtained with the access to a half of the state in fact follows from the normalization condition.

This result can be also obtained in a somewhat different way, which clarifies distinctive features of the situations when a part of the Hilbert space of states (i.e., a part of the carrier of state) is accessible for the measurement. Let us find the measurement minimizing the error in distinguishing two states described by the density matrices in the case when a part of the state carrier is accessible. The form of the density matrices is as follows:

$$\rho_{0,1} = \left\{\frac{1}{\sqrt{2}}\left(\int_{-\infty}^{\infty}[f(\tau) + f(\tau-\tau_0)]|\tau\rangle d\tau\right)\right.$$
$$\left.\times\ \frac{1}{\sqrt{2}}\left(\int_{-\infty}^{\infty}[f^*(\tau') + f^*(\tau'-\tau_0)]\langle\tau'|d\tau'\right)\right\} \quad (22)$$
$$\otimes\, |e_{0,1}\rangle\langle e_{0,1}| = \rho(f)\otimes\rho(0,1).$$

Now let us derive an expression for the error in distinguishing the $\rho_{0,1}$ states when only a part of the space–time is accessible for the measurement. The problem formally reduces to the situation when a region $\Delta(\tau)$ is accessible, and the remaining part of the space–time is inaccessible for the measurement. The measurement is described by the expansion of unity comprising two terms representing the unity in the subspaces generated by the basis set vectors $|\tau\rangle$ belonging to the interval $\Delta(\tau)$ and by the vectors from the inaccessible space–time region denoted as

$$\bar{\Delta}(\tau) = (-\infty, \infty)\backslash\Delta(\tau).$$

This yields

$$I\otimes\mathbf{C}^2 = I(\Delta\tau)\otimes\mathbf{C}^2 + I(\bar{\Delta}(\tau))\otimes\mathbf{C}^2$$
$$= \left(\int_{\Delta(\tau)}|\tau\rangle\langle\tau|d\tau\right)\otimes\mathbf{C}^2 + \left(\int_{\bar{\Delta}(\tau)}|\tau\rangle\langle\tau|d\tau\right)\otimes\mathbf{C}^2. \quad (23)$$

Let the state ρ_0 be presented for the measurement at an aprioric probability of π_0 and the state ρ_1, at a probability of π_1 ($\pi_0 + \pi_1 = 1$). Below we use $\pi_0 = \pi_1 = 1/2$ (the states are simply equiprobably guessed by Bob).

Since only a part of the space–time is accessible for the measurement (which automatically implies restricted access to the Hilbert space of states because the basis states are numbered in τ), the total error is a sum of two contributions. The first contribution $P_e(\bar{\Delta}(\tau))$ corresponds to cases when the measuring instrument on the Bob side fails to operate (for outcomes in the inaccessible region). The second contribution $P_e(\Delta(\tau))$ describes the error in distinguishing states for outcomes in the region accessible to Bob.

The provability that a transmitted state was not detected by Bob (whose instrument failed to operate) is

$$P(\bar{\Delta}(\tau)) = \mathrm{Tr}\{(\pi_0\rho_0 + \pi_1\rho_1)(I(\bar{\Delta}(\tau))\otimes\mathbf{C}^2)\}$$
$$= \pi_0\rho_0 + \pi_1\rho_1, \quad (24)$$
$$p_0 = p_1 = p,$$
$$p = \frac{1}{2}\int_{\bar{\Delta}(\tau)}[|f(\tau)|^2 + |f(\tau-\tau_0)|^2]d\tau.$$

Here, the probability of an outcome in the inaccessible region for the state ρ_0 with the given aprioric probability of presentation π_0 is

$$p_0 = \frac{\pi_0 p}{\pi_0 p + \pi_1 p} = \pi_0. \quad (25)$$

By the same token, for the state ρ_1 presented with the aprioric probability π_1, the probability of an outcome in the inaccessible region is

$$p_1 = \frac{\pi_1 p}{\pi_0 p + \pi_1 p} = \pi_1. \quad (26)$$

The probability of an error (misidentification), whereby the state ρ_0 was presented for the state ρ_1 transmitted (and vice versa), is

$$P_e(\bar{\Delta}(\tau)) = \pi_0 p_1 + \pi_1 p_0. \quad (27)$$

Let only one half of the states to fall in the accessible region (the carrier $f(\tau)$ or $f(\tau-\tau_0)$ occurs in the inaccessible region) and, accordingly, another half to be in the region inaccessible for the measurement. Then, according to (24)–(27), the probability of an error for the outcome in the inaccessible region is $P_e(\bar{\Delta}(\tau)) = 1/2$.

In the general case, the measurement minimizing the probability of an error is given (for a binary statistical decision function) by a unity expansion of the type

$$\tilde{E}_0 + \tilde{E}_1 = I(\Delta(\tau))\otimes I_{\mathbf{C}^2}$$
$$= I(\Delta(\tau))\otimes(E_0 + E_1), \quad (28)$$

where, in contrast to [34, 36], the expansion is limited to subspace $\Delta(\tau)$. The minimum probability of an error is determined by minimizing over all possible expansions (for detail, see [34, 36])

$$P_e(\Delta(\tau))$$
$$= \min_{\{\tilde{E}_0,\,\tilde{E}_1\}}(\pi_0\mathrm{Tr}\{\rho_0\tilde{E}_1\} + \pi_1\mathrm{Tr}\{\rho_1\tilde{E}_0\}), \quad (29)$$

where π_0 and π_1 are the probabilities of density matrices ρ_0 and ρ_1, respectively (in our example, $\pi_0 = \pi_1 = 1/2$ are the probabilities of preparing the states corresponding to 0 and 1).

Taking into account (28), an expression for the error probability can be written as

$$P_e(\Delta(\tau)) = \pi_0\mathrm{Tr}\{\rho(f)I(\Delta(\tau))\} + \mathrm{Tr}\{\Gamma\tilde{E}_0\},$$
$$\Gamma = \pi_1\rho_1 - \pi_0\rho_0. \quad (30)$$

Determination of the minimum of $P_e(\Delta(\tau))$ reduces to minimizing $\mathrm{Tr}\{\Gamma\tilde{E}_0\}$ over all possible operators $\tilde{E}_0$. Since the region accessible for the measurement is

reduced to the interval $\Delta(\tau)$ the trace is

$$\begin{aligned}\mathrm{Tr}\{\Gamma\tilde{E}_0\} &= \mathrm{Tr}_{\Delta(\tau)}\{\Gamma\tilde{E}_0\}\\ &= \mathrm{Tr}_{\Delta(\tau)}\{\rho(f)\}\mathrm{Tr}\{(\pi_1\rho(1)-\pi_0\rho(0))E_0\}\\ &= \left\{\frac{1}{2}\int\limits_{\Delta(\tau)}[|f(\tau)|^2+|f(\tau-\tau_0)|^2]d\tau\right\}\mathrm{Tr}\{\Gamma E_0\}.\end{aligned} \tag{31}$$

Taking into account that

$$0 \le E_0 \le I(\Delta(\tau)) \otimes I_{\mathbf{C}^2},$$

we obtain

$$\mathrm{Tr}\{\Gamma E_0\} \ge \mathrm{Tr}\{\Gamma\} = \sum_i \gamma_i. \tag{32}$$

The minimum error is determined by the negative eigenvalues γ_i of the operator

$$\Gamma = \pi_1\rho(1) - \pi_0\rho(0)$$

(see [36]). Operators $\tilde{E}_0$ must obey the conditions

$$\begin{aligned}\langle\gamma_i|\tilde{E}_0|\gamma_i\rangle &= 1, \quad \gamma_i \le 0,\\ \langle\gamma_i|\tilde{E}_0|\gamma_i\rangle &= 0, \quad \gamma_i \ge 0,\end{aligned} \tag{33}$$

where $|\gamma_i\rangle$ are eigenvectors of the operator

$$\Gamma = \sum_i \gamma_i|\gamma_i\rangle\langle\gamma_i|.$$

The matrix of Γ in the basis set $\{|e_0\rangle, |e_1\rangle\}$ has the form

$$\Gamma = \begin{pmatrix}\pi_1 & 0\\ 0 & -\pi_0\end{pmatrix}, \tag{34}$$

a single $\gamma_2 = -\pi_0 = -1/2$. Accordingly,

$$\begin{aligned}\tilde{E}_0 &= I(\Delta(\tau)) \otimes \begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix},\\ \tilde{E}_1 &= I(\Delta(\tau)) \otimes \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}.\end{aligned} \tag{35}$$

The minimum error is

$$P_e(\Delta(\tau)) = \overline{f^2(\Delta(\tau))}\left(\pi_0 + \sum_{\gamma_i \le 0}\gamma_i\right), \tag{36}$$

where

$$\overline{f^2(\Delta(\tau))} = \frac{1}{2}\int\limits_{\Delta(\tau)}[|f(\tau)|^2+|f(\tau-\tau_0)|^2]d\tau. \tag{37}$$

Finally, we obtain for the error probability corresponding to outcomes in the region accessible for the measurements

$$\begin{aligned}P_e(\Delta(\tau)) &= \overline{f^2(\Delta(\tau))}(\pi_0-\pi_0) = \frac{1}{2}\left(\frac{1}{2}-\frac{1}{2}\right) = 0,\\ \overline{f^2(\Delta(\tau))} &= \frac{1}{2}.\end{aligned} \tag{38}$$

On the intuitive level, this result can be interpreted as follows. Assume that we have to distinguish between two spatially extended one-photon states with different (orthogonal) helicities. The probability of distinguishing these states is unity only when the states are accessible as a whole (accordingly, the error probability is zero). Although the basis vectors for different helicity values are orthogonal, the states are identified with uncertainty due to their spatial extension (making the spatial part of the states inaccessible as a whole). From the physical standpoint, this is related to the fact that a helicity state cannot exist outside the internal degrees of freedom. Owing to the normalization with respect to the spatial degrees of freedom, the probability of the instrument operation on Bob's side does not exceed unity for any measurement. The unambiguous distinguishing between states even with different (orthogonal) helicities basically requires a finite time because limitations posed by the special relativity make it impossible to access the whole state within a time shorter than the effective extension of the state divided by the speed of light.

For a large number of trials, the total error equals the fraction of incorrectly identified states. The fraction (probability) of outcomes in the accessible region is

$$N(\Delta(\tau)) = \mathrm{Tr}\{(\pi_0\rho_0+\pi_1\rho_1)(I(\Delta(\tau)) \otimes I_{\mathbf{C}^2})\}, \tag{39}$$

and that in accessible region,

$$N(\bar{\Delta}(\tau)) = \mathrm{Tr}\{(\pi_0\rho_0+\pi_1\rho_1)(I(\bar{\Delta}(\tau)) \otimes I_{\mathbf{C}^2})\}. \tag{40}$$

The total error probability is a sum of the error in the inaccessible region multiplied by the fraction of such outcomes and the error in the accessible region multiplied by the corresponding fraction:

$$P_e = P_e(\bar{\Delta}(\tau))N(\bar{\Delta}(\tau)) + P_e(\Delta(\tau))N(\Delta(\tau)). \tag{41}$$

In the case of access to a half of the states, this yields

$$\begin{aligned}P_e &= \frac{1}{2}\times\frac{1}{2} + 0\times\frac{1}{2} = \frac{1}{4},\\ P_e(\bar{\Delta}(\tau)) &= \frac{1}{2}, \quad N(\bar{\Delta}(\tau)) = \frac{1}{2},\\ P_e(\Delta(\tau)) &= 0, \quad N(\Delta(\tau)) = \frac{1}{2}.\end{aligned} \tag{42}$$

Accordingly, the probability of the correct identification amounts to 3/4.

This result can be interpreted as follows. If Alice randomly and equiprobably prepares one of the states (ρ_0 or ρ_1) and presents it to Bob for the measurement, the measuring instrument operates with the probability 1/2 in one of the channels (corresponding to 0 or 1). Once the instrument has operated, the states are unambiguously identified. However, if the instrument fails to operate, Bob has to guess the state. The probability of this situation is 1/2 and the probability of simply guessing is 1/2. For such events, the probability of the correct identification is

$$1/2 \times 1/2 = 1/4.$$

The total probability of the correct identification is

$$1/2 + 1/4 = 3/4.$$

In this case, the probability $1 - P_e$ coincides with the probability of correctly identifying the secret bit. This probability (exceeding 1/2) of the correct identification is unacceptably high for constructing the protocol. The situation dramatically changes if the secret bit represents the bit of parity for N states. As will be shown below, the probability of the correct identification of the parity bit by Bob in this case exceeds the value 1/2 (the probability of simple guessing that represents the worst variant) only by an exponentially small quantity.

Now let us consider the probability of an error in distinguishing between 0 and 1 when these values are encoded as the parity bit determined over N orthogonal states. The problem of determining the parity bit over N states, with every bit being encoded by a nonorthogonal state and the space of states being accessible as a whole, was considered previously by Bennett *et al.* [37].

Let us first calculate the error of distinguishing the states for outcomes in the accessible region. For an N-bit line (corresponding to a density matrix $\rho_{0,1}$) randomly selected from 2^N possible combinations (among which $2^N/2$ are even and $2^N/2$ are odd), the problem reduces to distinguishing two density matrices corresponding to the even and odd lines:

$$\begin{aligned}\hat{\rho}_0 &= \frac{2}{2^N}\sum_{(i_1 \oplus i_2 \oplus \ldots \oplus i_N) = 0} \overbrace{\rho_{i_1} \otimes \rho_{i_2} \otimes \ldots \otimes \rho_{i_N}}^{N} \\ &= \frac{2}{2^N}(\rho(f) \otimes \rho(f) \otimes \ldots \otimes \rho(f)) \\ &\otimes \sum_{(i_1 \oplus i_2 \oplus \ldots \oplus i_N) = 0} \rho(i_1) \otimes \rho(i_2) \otimes \ldots \otimes \rho(i_N), \\ &\quad i_k = 0, 1, \quad k = 1, \ldots, N.\end{aligned} \tag{43}$$

$$\begin{aligned}\hat{\rho}_1 &= \frac{2}{2^N}\sum_{(i_1 \oplus i_2 \oplus \ldots \oplus i_N) = 1} \overbrace{\rho_{i_1} \otimes \rho_{i_2} \otimes \ldots \otimes \rho_{i_N}}^{N} \\ &= \frac{2}{2^N}(\rho(f) \otimes \rho(f) \otimes \ldots \otimes \rho(f)) \\ &\otimes \sum_{(i_1 \oplus i_2 \oplus \ldots \oplus i_N) = 1} \rho(i_1) \otimes \rho(i_2) \otimes \ldots \otimes \rho(i_N), \\ &\quad i_k = 0, 1, \quad k = 1, \ldots, N.\end{aligned} \tag{44}$$

The measurement minimizing the error in distinguishing the density matrices $\hat{\rho}_0$ and $\hat{\rho}_1$ is described by the unity expansion

$$\begin{aligned}(I(\Delta(\tau)) \otimes I_{\mathbf{C}^2})^{\otimes N} &= \hat{\tilde{E}}_0 + \hat{\tilde{E}}_1 \\ &= I(\Delta(\tau))^{\otimes N} \otimes (\hat{E}_0 + \hat{E}_1), \\ \hat{E}_0 + \hat{E}_1 &= I_{\mathbf{C}^2}^{\otimes N}.\end{aligned} \tag{45}$$

The corresponding error probability is

$$\begin{aligned}P_e(\Delta(\tau)) &= \pi_0 \mathrm{Tr}\left\{\hat{\rho}_0 (I(\Delta(\tau))^{\otimes N} \otimes I_{\mathbf{C}^2}^{\otimes N})\right\} \\ &+ \mathrm{Tr}\{\hat{\Gamma}(I(\Delta(\tau))^{\otimes N} \otimes \hat{E}_0)\}.\end{aligned} \tag{46}$$

Accordingly, the minimum error is given by a formula analogous to (28) and determined by the negative eigenvalues γ_i of the operator

$$\hat{\Gamma} = (\overline{f^2(\Delta)})^N(\pi_1\hat{\rho}(1) - \pi_0\hat{\rho}(0)) = (\overline{f^2(\Delta)})^N\Gamma. \tag{47}$$

In the basis of vectors ordered into even and odd sets with respect to the sum of indices, the operator Γ has the following form:

$$\Gamma = \frac{1}{2^N}\begin{pmatrix} 1 & 0 & 0 & \ldots \\ 0 & 1 & 0 & \ldots \\ . & . & . & \ldots \\ . & . & . & \ldots \\ 0 & \ldots & -1 & 0 \\ 0 & 0 & \ldots & -1 \end{pmatrix},$$

$$\text{even} \begin{cases} |e_0\rangle\ldots|e_1\rangle \\ \quad \ldots \\ |e_1\rangle\ldots|e_0\rangle, \end{cases} \tag{48}$$

$$\text{odd} \begin{cases} |e_0\rangle\ldots|e_0\rangle \\ \quad \ldots \\ |e_1\rangle\ldots|e_1\rangle. \end{cases}$$

Finally, the minimum probability of an error in determining the resulting parity bit over N orthogonal states inaccessible for measurement as a whole (for $\pi_0 = 1/2$) is

$$P_e(\Delta(\tau)) = (\overline{f^2(\Delta)})^N\left(\frac{1}{2} + \left(\frac{1}{2^N}\sum_{\gamma_i \le 0}^{2^N/2}(-1)\right)\right). \quad (49)$$

In the case of access to a half of each state ($\overline{f^2(\Delta)} = 1/2$), we obtain

$$P_e(\Delta(\tau)) = \left(\frac{1}{2}\right)^N\left(\frac{1}{2} - \frac{1}{2^N}\frac{2^N}{2}\right) = 0. \quad (50)$$

Due to the orthogonality of the channels, the error probability (for the outcome in the accessible region) is zero. This formula has to be interpreted as follows: if all N outcomes took place in the accessible region, the error probability is zero due to orthogonality of the channels; the same is valid if the outcome in the accessible region took place for m states (in this case, m should be substituted for N in (47) and (48)). Thus, if the outcomes related to the presented states took place in the accessible region, the corresponding states become unambiguously identified.

However, the outcomes can also take place in the region inaccessible for the measurements. As a state propagating into the accessible region becomes accessible as a whole ($\overline{f^2(\Delta)} \longrightarrow 1$), the error probability tends to zero ($P_e \longrightarrow 0$). The orthogonal states are unambiguously distinguishable when accessible as a whole.

Now let us evaluate the probability of correctly identifying the parity bit. The total number of N-bit binary lines is 2^N and the outcomes may take place in both accessible and inaccessible regions. The space of events contains two elements. The first represents the case when all N outcomes take place in the accessible region. In this case, the probability of correctly identifying the parity bit is unity. However, the probability of this event (for only a half of each state being accessible) is

$$\overline{f^2(\Delta)}^N = 2^{-N}.$$

The second element of the space of events represents all other cases (i.e., those with no more that $N - 1$ outcomes in the accessible region). The probability of all these outcomes (for only a half of each state being accessible for the measurement) is

$$\sum_{k=0}^{N-1} C_N^k \overline{f^2(\Delta)}^k (1 - \overline{f^2(\Delta)})^{N-k} = \sum_{k=0}^{N-1} C_N^k \frac{1}{2^k}\frac{1}{2^{N-k}} = 1 - \overline{f^2(\Delta)}^N = 1 - 2^{-N}. \quad (51)$$

For such outcomes, the probability of an error in determining the parity bit is 1/2. Indeed, Bob knows a k-bit line ($k \le N - 1$), the parity of which is unambiguously established. However, the remaining $(N - k)$-bit line (for outcomes in the inaccessible region) can be either even or odd with a probability of 1/2. Therefore, the resulting parity of the total N-bit line is also known only at a probability of 1/2, since the knowledge of the parity of a part of the k-bit line does not influence the probability of correctly determining the parity of the whole line.

The resulting error of determination of the parity bit is a sum of two contributions. The first contribution is due to the event corresponding to all outcomes taking place in the accessible region, and the second reflects all other events. Each of these terms is the product of the probability of an error in determining the parity bit by the probability of the corresponding event. Eventually, this yields

$$P_e(\text{parity}) = \frac{1}{2}(1 - 2^{-N}) + 0 \times 2^{-N} = \frac{1}{2} - \frac{1}{2} \times 2^{-N}. \quad (52)$$

Accordingly, the probability that Bob correctly identifies the parity bit (for the access to half of the states) is

$$P_c(\text{parity}) = 1 - P_e(\text{parity}) = \frac{1}{2} + \frac{1}{2} \times 2^{-N} \quad (53)$$

which exceeds the probability of simple guessing only by an exponentially small quantity.

Thus, within a period of time τ_0 ($\Delta\tau < \tau < \Delta\tau + \tau_0$) after beginning of the protocol, Bob possesses an exponentially small amount of information about the secret bit.

However, the scheme of the secret bit presentation in the form of a parity bit for an N-bit line is insufficient for a secure protocol, since it leaves an unacceptably large possibility for Alice to deceive Bob (a large probability of undisclosed delay in the bit selection). In order to eliminate this possibility, it is necessary to encode each of the N bits with a block of k like bits (the k value is determined below), which are sent to Bob intermittently via Nk channels.

Finally, let us present an expression for the probability of an error in distinguishing the density matrices corresponding to 0 and 1 in the case when the parity bit is encoded by k-bit blocks (containing either all zeros or all unities). According to the protocol, the secret bit selected by Alice is a parity bit of N zeros or unities, with each 0 and 1 presented by blocks of k zeros or unities. This block representation of each bit is necessary to enable Bob monitoring the absence of deception on Alice's side.

In this case, the total number of possible binary line combinations is 2^{Nk}. The number of even and odd com-

binations (encoded by k-bit blocks of all zeros or unities) is

$$S_{\text{odd, even}} = \frac{1}{2}\sum_{m=0}^{N-1} C_{Nk}^{mk} = 2^{Nk}\left(\frac{1}{2k}\right)\sum_{l=1}^{k}\cos^{Nk}\left(\frac{l\pi}{k}\right)\cos(Nl\pi) \approx 2^{Nk}, \tag{54}$$

which equals in fact to the total number of arrangements for $(N-l)k$ unities and lk zeros ($0 \le l \le N$) over Nk cells [35].

Note that, if the position of each k-block were fixed (with zeros and unities of different blocks not mixed), the number of possible combinations of the even and odd blocks would be only 2^N which is exponentially smaller than 2^{Nk} for large k values (see formula (54)).

By definition, the rest of the combinations numbered

$$N_{\text{rest}} = 2^{Nk} - S_{\text{odd}} - S_{\text{even}} \ll 2^{Nk}$$

enter neither S_{odd} nor S_{even} in the form of any subdivisions into k-blocks (containing either all zeros or all unities). This circumstance is important for the proposed protocol.

In the bass set ordered for the even and odd k-block states and the rest of the states (for certainty, k is assumed even)

$$S_{\text{even}} \longrightarrow \begin{cases} \overbrace{|e_0\rangle \otimes |e_0\rangle \otimes \ldots \otimes |e_0\rangle}^{k} \otimes \overbrace{\ldots\ldots}^{k\ \ldots\ k} \otimes \overbrace{|e_0\rangle \otimes |e_0\rangle \otimes \ldots \otimes |e_0\rangle}^{k} \\ \text{other combinations of } |e_0\rangle \text{ and } |e_1\rangle \text{ with even} \\ \text{numbers of } k\text{-blocks } \overbrace{|e_1\rangle \otimes |e_1\rangle \otimes \ldots \otimes |e_1\rangle}^{k} \\ \text{(a total of } N \text{ blocks)} \\ \overbrace{|e_1\rangle \otimes |e_1\rangle \otimes \ldots \otimes |e_1\rangle}^{k} \otimes \overbrace{\ldots\ldots}^{k\ \ldots\ k} \otimes \overbrace{|e_1\rangle \otimes |e_1\rangle \otimes \ldots \otimes |e_1\rangle}^{k} \end{cases} \tag{55}$$

$$S_{\text{odd}} \longrightarrow \begin{cases} \overbrace{|e_0\rangle \otimes |e_0\rangle \otimes \ldots \otimes |e_0\rangle}^{k} \otimes \overbrace{\ldots\ldots}^{k\ \ldots\ k} \otimes \overbrace{|e_1\rangle \otimes |e_1\rangle \otimes \ldots \otimes |e_1\rangle}^{k} \\ \text{other combinations of } |e_0\rangle \text{ and } |e_1\rangle \text{ with odd} \\ \text{numbers of } k\text{-blocks } \overbrace{|e_1\rangle \otimes |e_1\rangle \otimes \ldots \otimes |e_1\rangle}^{k} \\ \text{(a total of } N \text{ blocks)} \\ \overbrace{|e_1\rangle \otimes |e_1\rangle \otimes \ldots \otimes |e_1\rangle}^{k} \otimes \overbrace{\ldots\ldots}^{k\ \ldots\ k} \otimes \overbrace{|e_0\rangle \otimes |e_0\rangle \otimes \ldots \otimes |e_0\rangle}^{k} \end{cases} \tag{56}$$

the operator $\hat{\Gamma}$ analogous to that in (48) has the following form:

$$\hat{\Gamma} = \begin{pmatrix} \hat{I}_{S_{\text{odd}}} & 0 & 0 \\ 0 & -\hat{I}_{S_{\text{even}}} & 0 \\ 0 & 0 & \hat{0} \end{pmatrix}, \tag{57}$$

where $\hat{I}_{S_{\text{odd}}}$ ($\hat{I}_{S_{\text{even}}}$) are the identity matrices of the size $S_{\text{odd}} \times S_{\text{odd}}$ ($S_{\text{even}} \times S_{\text{even}}$) and $\hat{0}$ is the zero matrix of the size $N_{\text{rest}} \times N_{\text{rest}}$.

In the same basis set, the measurement operators $\hat{E}_0$ and $\hat{E}_1$ are and follows:

$$\hat{E}_0 = \begin{pmatrix} \hat{0} & 0 & 0 \\ 0 & \hat{I}_{S_{\text{even}}} & 0 \\ 0 & 0 & \hat{0} \end{pmatrix}, \quad \hat{E}_1 = \begin{pmatrix} \hat{I}_{S_{\text{odd}}} & 0 & 0 \\ 0 & \hat{0} & 0 \\ 0 & 0 & \hat{I} \end{pmatrix}. \tag{58}$$

The probability of an error in determining the k-block parity bit for Nk outcomes in the accessible region is

$$P_e(\Delta(\tau)) = \left(\overline{f^2(\Delta)}\right)^{Nk} \times \left(\frac{1}{2} - \frac{1}{2^{(S_{\text{even}} + S_{\text{odd}})}}\sum_{\substack{i=1 \\ \gamma_i \le 0}}^{2^{S_{\text{even}}}} (-1)\right) = 0. \tag{59}$$

The error is zero for any number of states, the measurements on which gave outcomes in the accessible region.

Now we will determine the error of identification of the parity bit for the k-block representation. Note that the outcomes may take place in both accessible and inaccessible regions. Let us first calculate the minimum necessary number of outcomes in the accessible region for which the line parity is unambiguously identified. Since the direct calculation is rather cumbersome, we will obtain an estimate using the following considerations (essentially, in terms of the Shannon typical sequences [38]; see also [39]). In the situation with every bit representing a block of unit length ($k = 1$), the power of the set of all lines is $\Omega = 2^N$. Accordingly, the information from each element of the set is $I = \log_2|\Omega|$, representing (to within a rounding error) the number of binary symbols necessary for individualization of the elements. If each symbol (in our case, detector operation in the accessible region) appears at a probability of p, the probability of the element identification is p^I.

In the k-block representation, the subset power is given by expression (54) and the number of binary symbols necessary for individualization of the elements is

$$I = \log_2\left(\frac{2^{Nk-1}}{k}\sum_{l=1}^{k}\cos^{Nk}\left(\frac{l\pi}{k}\right)\cos(Nl\pi)\right) = \alpha(N,k)(Nk). \quad (60)$$

This is the number of outcomes in the accessible region, which is necessary for identification of the line parity. The probability of such an event (for $p = \overline{f^2(\Delta)} = 1/2$) is

$$P_{\text{acc}} = p^{\alpha(N,k)(Nk)} = 2^{-\alpha(N,k)(Nk)}. \quad (61)$$

For these outcomes, the error is zero. The probability of outcomes in the inaccessible region is

$$P_{\text{unacc}} = 1 - P_{\text{acc}} = 1 - 2^{-\alpha(N,k)(Nk)}; \quad (62)$$

and the corresponding error in determining the parity bit amounts to

$$P_e(\text{parity}) = \frac{1}{2}(1 - 2^{-\alpha(N,k)(Nk)}) + 0 \times 2^{-\alpha(N,k)(Nk)}. \quad (63)$$

Thus, the probability of the correct identification of the parity bit in this case exceeds 1/2 (the probability of simple guessing) only by an exponentially small quantity:

$$P_c(\text{parity}) = 1 - P_e(\text{parity}) = \frac{1}{2} + 2^{-\alpha(N,k)(Nk)}. \quad (64)$$

It should be noted that the number of lines in the block representation with zeros and unities from various blocks intermixed equals in the order of magnitude the total number of lines ($\approx 2^{Nk}$); each k-block line appears almost as if the block length is 1. For the line parity determination, almost the whole line is necessary (to within a correction factor $\alpha(N, k)$ (see formula (60)).

If the block positions were fixed, the set power would be equal to $2^N/2$ and N binary tests would be necessary. However, the probability p of each test equals a sum of the probabilities of having 1, 2, …, or k outcomes in the accessible region for each block:

$$p = \sum_{l=1}^{k} C_k^l \frac{1}{2^l}\frac{1}{2^{(k-l)}} = 1 - 2^{-k}. \quad (65)$$

The probability of N outcomes in the accessible region (allowing Bob to know the parity on having the access to halves of the states)

$$P_{\text{acc}} = p^N = (1 - 2^{-k})^N \quad (66)$$

would be high for comparable N and k.

Now we have only to show that, as the whole states become accessible for the measurement after elapse of the time $\tau_0 + \Delta\tau \approx \tau_0$, the probability for Bob being deceived by Alice tends to zero. Strictly speaking, it is necessary to show that Alice cannot change the secret bit selection once the protocol has been started (in other words, the probability that Alice changes the selection after protocol start and remains undisclosed is exponentially small 2^{-Nk}).

According to the protocol, Nk states are simultaneously sent to Bob intermittently via Nk channels. A check for the absence of deception on the Alice side is provided by a measurement described by a unity expansion of the type

$$I^{\otimes^{Nk}} \otimes I_{\mathbf{C}^2}^{\otimes^{Nk}} = (\mathcal{P}_0(f) + \mathcal{P}_1(f) + \mathcal{P}_\perp)^{\otimes^{Nk}}, \quad (67)$$

where

$$\mathcal{P}_{0,1}(f) = \left(\frac{1}{\sqrt{2}}\int_{-\infty}^{\infty}[f(\tau) + f(\tau - \tau_0)]|\tau\rangle d\tau\right) \times \left(\frac{1}{\sqrt{2}}\int_{-\infty}^{\infty}[f^*(\tau') + f^*(\tau' - \tau_0)]\langle\tau'|d\tau'\right) \otimes |e_{0,1}\rangle\langle e_{0,1}|, \quad (68)$$

$$\mathcal{P}_\perp = I \otimes I_{\mathbf{C}^2} - \mathcal{P}_0(f) - \mathcal{P}_1(f).$$

In each of the Nk measurement channels, there are three possible outcomes corresponding to $\mathcal{P}_0(f)$, $\mathcal{P}_1(f)$, and $\mathcal{P}_\perp(f)$. If Alice sends true states, the probabilities of outcomes are as follows:

$$\begin{aligned}
&\Pr\{\rho_0, 0\} \equiv \mathrm{Tr}\{\rho_0\mathcal{P}_0(f)\} \equiv 1,\\
&\Pr\{\rho_1, 1\} \equiv \mathrm{Tr}\{\rho_1\mathcal{P}_1(f)\} \equiv 1,\\
&\Pr\{\rho_0, 1\} \equiv \mathrm{Tr}\{\rho_0\mathcal{P}_1(f)\} \equiv 0,\\
&\Pr\{\rho_1, 0\} \equiv \mathrm{Tr}\{\rho_1\mathcal{P}_0(f)\} \equiv 0,\\
&\Pr\{\rho_{0,1}, \perp\} \equiv \mathrm{Tr}\{\rho_{0,1}\mathcal{P}_\perp(f)\} \equiv 0.
\end{aligned} \quad (69)$$

Thus, when Alice sends true states, all outcomes with a probability of unity must take place only in $\mathcal{P}_0(f)$, $\mathcal{P}_1(f)$.

A delay in the state selection by Alice for a time exceeding $2\Delta\tau$ would imply that the states have to be used not covering the first halves of the true extended states. In such cases, Alice begins to prepare the state after the protocol starts, with a delay exceeding $2\Delta\tau$. In all such states ρ, the carriers of which do not cover the first halves of true states, the outcome in the $\mathcal{P}_0(f)$, $\mathcal{P}_1(f)$ channels does not exceed 1/2. Indeed,

$$\begin{aligned}\mathrm{Tr}\{\rho\mathcal{P}_{0,1}(f)\} &= \frac{1}{2}\int_{-\Delta\tau}^{\Delta\tau}\int_{-\Delta\tau}^{\Delta\tau} f(\tau)\rho(\tau,\tau')f^*(\tau')d\tau d\tau' \\ &+ \frac{1}{2}\int_{\tau_0-\Delta\tau}^{\tau_0+\Delta\tau}\int_{\tau_0-\Delta\tau}^{\tau_0+\Delta\tau} f(\tau)\rho(\tau,\tau')f^*(\tau')d\tau d\tau' \\ &\le \frac{1}{2}\frac{1}{(2\Delta\tau)^2}\int_{-\Delta\tau}^{\Delta\tau}\int_{-\Delta\tau}^{\Delta\tau}|\rho(\tau,\tau')|d\tau d\tau' \le \frac{1}{2}, \\ &|\rho(\tau,\tau')| \le 1,\end{aligned} \tag{70}$$

if the carrier

$$\rho = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\rho(\tau,\tau')|\tau\rangle\langle\tau'|d\tau d\tau',$$

$$\mathrm{Tr}\{\rho\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta_+(\tau-\tau')\rho(\tau,\tau')d\tau d\tau' = \int_{-\infty}^{\infty}\rho(\tau,\tau)d\tau = 1,$$

does not cover the first half of the true state:

$$\mathrm{supp}\rho(\tau,\tau') \cap \mathrm{supp}\, f(\tau-\tau_0) = \backslash\,\mathrm{varnothing}.$$

Thus, for the ideal communication channel, any delay in the state selection results in that the outcome probability in $\mathcal{P}_0(f)$, $\mathcal{P}_1(f)$ does not exceed 1/2. More strictly speaking, in each single experiment, the outcome of measurements even on the states delayed by more than $2\Delta\tau$ may take place only in $\mathcal{P}_0(f)$, $\mathcal{P}_1(f)$ and be absent in $\mathcal{P}_\perp(f)$ with a probability of 1/2. However, for k tests, the probability that all outcomes of the measurements take place in $\mathcal{P}_0(f)$, $\mathcal{P}_1(f)$ and are absent in $\mathcal{P}_\perp(f)$, thus simulating the statistics for undelayed states, is as small as 2^{-k}. This circumstance is employed in the protocol below.

Now we will proceed with formulating the protocols.

3. BIT COMMITMENT PROTOCOL FOR THE STATES WITH COMPACT CARRIER IN THE IDEAL COMMUNICATION CHANNEL

1. Prior to proceeding with the protocol, the parties agree on the type of states (localization $\Delta\tau$ and the form of the state carrier $f(\tau)$) and the protocol duration during which Alice keeps the secret bit. The latter time can be selected arbitrarily large (not considering related technical difficulties). Alice and Bob also preliminarily agree in the N and k values.

2. Alice selects a secret bit representing a bit of parity for a line of N representatives

$$b = \sum_{i=1}^{N} a[i,j],$$

where $a[i,j]$ are bits 0 or 1, representing jth block (j is the number of a k-block); all bits in one block (with the same number j) are alike.

3. At the protocol onset moment (which is also agreed preliminarily), Alice begins with preparing Nk extended states, each of two halves, which propagate in Nk quantum communication channels. The states can also be sent sequentially via the same channel, which would only increase the protocol duration. The channel length is assumed to be zero (with an allowance for the above remarks). This in fact implies that Bob controls only his laboratory (the vicinity of point x_B) and can monitor neither the rest of space nor the communication channel, that is, Alice may reside immediately at Bob's laboratory door. Alice can control only the vicinity of point x_A where the states are prepared. The states of various k-blocks $a[i,j]$ are sent intermittently via different channels.

4. A time for the unveiling stage can be selected by Bob at any instant within the interval $\Delta\tau < \tau < \tau_0 + \Delta\tau$. By Bob's request, Alice must communicate via a classical channel which states were transmitted via each quantum channel and assign each channel to a certain block (i.e., indicate the channel and the representative value $a[i,j]$).

5. Bob performs measurements described by a unity expansion of the type (67), (68). The states are orthogonal (and, hence, unambiguously distinguishable), but the projectors $\mathcal{P}_{0,1}$ are nonlocal. For this reason, the unambiguous result for true states can be obtained by measurement (67) only provided access to all states as a whole, which requires the time $2\Delta\tau + \tau_0$. Then Bob compares the results of his measurements in each quantum channel to the data communicated by Alice via the classical channel. In the case of ideal quantum communication channels, all measurements in the channels belonging to the same block must give the results only in like channels (0 or 1). If Bob founds a discrepancy at least in one position between his results and the data disclosed by Alice, the protocol is terminated.

Note that, if Alice plays fairly (i.e., the parity of an Nk-bit line is actually selected and true extended states are sent at the protocol onset time), the parity bit cannot be changed through any rearrangements. Otherwise, we have to conclude that the sets of even and odd lines in the k-block representation may overlap.

6. If Alice fails to take a decision on the bit selection at the beginning of the protocol (or, strictly speaking, if she takes a decision upon elapse of the time interval $\Delta\tau$ but, naturally, before the exchange via the classical channel, i.e., within $\Delta\tau < \tau < \tau_0 + \Delta\tau$), she will send the states different from $|\psi_{0,1}\rangle$. The counts in each channel $\mathcal{P}_{0,1}$ for any states different from true would occur at a probability not exceeding 1/2. In order to change a decision on the secret bit, it is sufficient for Alice to delay a decision in one of the blocks, for which purpose she has to delay the states in a block of k bits. According to (70), the probability that Alice would remain undisclosed on sending k delayed states is 2^{-k}.

7. For Bob, the probability to obtain unambiguous information on the secret bit before having access to the whole states does not exceed $1/2 + 2^{-\alpha(N, k)Nk}$ (see formula (64)).

Thus, the above protocol realizes the basic idea of the bit commitment scheme, whereby one party transmits a part of the information (from which another party can obtain only an exponentially small information about the secret bit before the unveiling stage). At the same time, Alice cannot change the selected secret bit after onset of the protocol. More precisely, the probability of undisclosed change of the secret bit after the protocol start is exponentially small.

The proposed scheme allows a fair-play protocol to be realized at a probability of not worse than $1 - 2^{-k}$, which is exponentially close to unity for sufficiently large k.

4. COIN TOSSING PROTOCOL FOR THE STATES WITH COMPACT CARRIER IN THE IDEAL COMMUNICATION CHANNEL

Although a coin tossing protocol can be constructed based on the bit commitment scheme, it is expedient to formulate the protocol explicitly.

1. Alice and Bob agree on the states as in the protocol described above. At the agreed onset time, each of the two parties sends to anotherintermittently N blocks of k states, so that the parity bits over N blocks selected by Alice and Bob are b_A and b_B, respectively. It is preliminarily agreed who wins in cases when the resulting parity bit $b = b_A \oplus b_B$ is zero or unity.

2. At a certain time instant τ ($-\Delta\tau < \tau < \tau_0 + \Delta\tau$) one of the two parties (e.g., Alice) communicates via the classical channel to the other party (Bob) the numbers of messages and the assignment of messages to blocks (the information is communicated only for half of the blocks for all states). Then Bob communicates to Alice (only after receiving her data) analogous information for the other half of the channel numbers (not coinciding with the channel numbers communicated by Alice). Upon receiving this information from Bob, Alice communicates him the assignment to blocks for the remaining channel numbers and indicates the states sent in each of the blocks. Finally, Bob communicates to Alice the analogous information concerning the remaining channel numbers. Since the channel length is $\tau_{ch} < \tau_0$, the exchange via the classical channel can be carried out during the period of time when both parties have access to only half of each state.

3. As long as the whole states are inaccessible, the probability for each party to obtain information concerning the parity bit of the other party from the results of its own measurements by analogy with the previous protocol does not exceed 1/2.

4. The parties perform measurements described by a unity expansion of the type (67), (68). The states are orthogonal, but the projectors $\mathcal{P}_{0,1}$ are nonlocal. For this reason, the unambiguous result for true states can be obtained by measurement (69) only provided access to all states as a whole, which requires the time $2\Delta\tau + \tau_0$.

5. After elapse of the time $\tau_0 + \Delta\tau$, when the whole states become accessible to both parties, Alice and Bob check the statistics of measurements and the correspondence of the classical information exchange to the results of measurements in both channels. The protocol is terminated if there is a lack of agreement between the results of measurements and classical information (change from 0 to 1 or vice versa) in at least one of the channels.

6. The probability for each party to know the parity bit of the other party before having access to the whole states exceeds the probability of simple guessing by no more than an exponentially small quantity $2^{-\alpha(N, k)Nk}$ (this is analogous to the previous protocol). Eventually, a true parity bit (lot) $b = b_A \oplus b_B$ is established at a probability close unity $(1 - 2^{-k})$.

Obviously, one of the two parties (even if both are sending true states) can terminate the protocol in case of unfavorable resulting bit, pleading the lack of correspondence between the data obtained by classical exchange and quantum measurements. However, this situation falls outside the problem under consideration and has to be solved by different means.

It should be noted that the data exchange between parties via the classical channel is necessary in order to avoid the deceiving strategy, whereby one party retransmits back the quantum states of another party. In particular, one party may only "reflect" the states sent by another party, rather than transmit its own states. Using this strategy, the party preliminarily agreed to win when the total parity bit is

$$b = b_A \oplus b_B = 0,$$

may permanently win by deceiving another party, since $b_A \equiv b_B$ and $b = 0$.

The classical exchange with information about only half of the states is also necessary for eliminating the retransmission strategy. If one of the parties communicates data about all states, the second party using the strategy of retransmitting quantum states may also send this information (once it has became preliminarily known) back via the classical channel because $\tau_{ch} < \tau_0$. When information about half of the states is sequentially communicated, this strategy fails to win.

5. BIT COMMITMENT PROTOCOL FOR THE STATES WITH NONLOCAL CARRIER IN THE IDEAL COMMUNICATION CHANNEL

The protocols considered above referred to the states with compact carriers (functions $f(\tau) \in \mathcal{D}(\tau)$). The set of such functions forms a dense set in the space of functions describing states of the free field (functions $f(\tau) \in \mathcal{J}(\tau)$ never turn zero). However, the field theory does not forbid the states on a mass surface which are localized arbitrarily strongly and decrease at a rate arbitrarily close to exponential, for example (14). Therefore, the states can be always selected so that measurements within a finite region on the light cone τ would yield an outcome probability arbitrarily close to unity (i.e., with arbitrarily small contributions from the tails of states in the infinity). More strictly speaking, the states (functions $f(\tau)$) and the region of measurement can be selected so that the probability of obtaining the result in the region $\Delta(\tau)$ would be

$$\Pr\{\Delta(\tau); i, i\} = \mathrm{Tr}\left\{\left(\left(\int_{-\Delta\tau}^{\Delta\tau} \mathcal{M}(d\tau)\right) \otimes \mathcal{P}_i\right)|\psi_i\rangle\langle\psi_i|\right\}$$
$$= \int_{-\Delta\tau}^{\Delta\tau} |f(\tau)|^2 d\tau = 1 - e^{-\xi} \longrightarrow 1, \qquad (71)$$
$$|\psi_i\rangle = \int_{-\infty}^{\infty} f(\tau)|\tau\rangle d\tau \otimes |e_i\rangle, \quad f(\tau) \in \mathcal{J}(\tau), \quad i = 0, 1,$$

where ξ can be arbitrarily large. The contribution from the tails of states outside the region $(-\Delta\tau, \Delta\tau)$ is

$$e^{-\xi} = \int_{|\tau| > \Delta\tau} |f(\tau)|^2 d\tau$$
$$= \mathrm{Tr}\left\{\left(\left(\int_{|\tau| > \Delta\tau} \mathcal{M}(d\tau)\right) \otimes \mathcal{P}_i\right)|\psi_i\rangle\langle\psi_i|\right\}. \qquad (72)$$

The case of states with a compact carrier, the extended states with nonlocal carrier will be written in the following form:

$$|\psi_{0,1}\rangle = \frac{1}{\sqrt{2}}\int_{-\infty}^{\infty} [f(\tau) + f(\tau - \tau_0)]|\tau\rangle d\tau \otimes |e_{0,1}\rangle, \qquad (73)$$
$$f(\tau) \in \mathcal{J}(\tau),$$

where the carrier $f(\tau)$ or $f(\tau - \tau_0)$ is strongly localized (similarly to that for the single-hump state (13) considered above) in the interval $(-\Delta\tau, \Delta\tau)$ or $(-\Delta\tau + \tau_0, \Delta\tau + \tau_0)$, respectively. The normalization condition yields

$$\frac{1}{2}\int_{-\Delta\tau}^{\Delta\tau} |f(\tau)|^2 d\tau = \frac{1}{2}\int_{-\Delta\tau+\tau_0}^{\Delta\tau+\tau_0} |f(\tau - \tau_0)|^2 d\tau$$
$$= \frac{1}{2} - \frac{1}{2}e^{-\xi} \longrightarrow \frac{1}{2} \qquad (74)$$

and

$$\frac{1}{2}\int_{\tau > |\Delta\tau|} |f(\tau)|^2 d\tau + \frac{1}{2}\int_{\tau + \tau_0 > |\Delta\tau|} |f(\tau - \tau_0)|^2 d\tau$$
$$+ \frac{1}{2}\int_{-\infty}^{\infty} [f^*(\tau) f(\tau - \tau_0) + f(\tau) f^*(\tau - \tau_0)] d\tau = e^{-\xi}. \qquad (75)$$

The measurement on the extended state in a finite window $\Delta(\tau_0) = (-\Delta\tau, \tau_0 + \Delta\tau)$ yields the result at a probability

$$\Pr\{\Delta(\tau_0); i, i\}$$
$$= \mathrm{Tr}\left\{\left(\left(\int_{-\Delta\tau}^{\Delta\tau+\tau_0} \mathcal{M}(d\tau)\right) \otimes \mathcal{P}_{0,1}\right)|\psi_{0,1}\rangle\langle\psi_{0,1}|\right\} \qquad (76)$$
$$= 1 - O(e^{-\xi}).$$

The last term (not exceeding $O(e^{-\xi})$) arises due to the overlap of tails of the state halves centered at $\tau = 0$ and $\tau = \tau_0$.

Thus, the statistics of measurements on the extended states must yield the results in the interval $(-\Delta\tau, \Delta\tau + \tau_0)$ with a probability $1 - O(e^{-\xi}) \longrightarrow 1$ (exponentially close to unity). Outside this interval, the probability of counts does not exceed $O(e^{-\xi})$ and can be made arbitrarily small by properly selecting $f(\tau)$, $\Delta\tau$, and τ_0.

Preparing a delocalized state with $f(\tau) \in \mathcal{J}(\tau)$ formally requires either an infinite time (e.g., if the state is generated by a point source) or access to the whole coordinate space (if the state is prepared at a given time instant by a nonlocal source). Any real protocol may only last for a finite time. In order to avoid such formal difficulties, a convenient (and common) practice is to consider the situation as follows. Alice controls the vicinity of point x_A and adiabatically ($t \longrightarrow -\infty$) switches on the source generating a vector $|\psi_{0,1}(\tau_0)\rangle$ from the vacuum state. The source operation is described (not dwelling on the particular realization of

such a source) by the $\hat{S}(\tau, -\infty)$ matrix acting upon the vacuum state:

$$|\psi_{0,1}(\tau)\rangle = \hat{S}(\tau, -\infty)|0\rangle = \int_{-\infty}^{\tau} [f(\tau') + f(\tau' - \tau_0)]|\tau'\rangle d\tau' \otimes |e_{0,1}\rangle. \quad (77)$$

The generated state is submitted to the communication channel.

On the intuitive level, this source can be considered as an atomic system (e.g., an atom) with an appropriate spectrum, excited by the adiabatically switched classical source of a special shape. The excitation is radiated into the communication channel (for the preparation of one- and two-photon states see, e.g., [40]).

Bob performs measurements described by a unity expansion analogous to (67):

$$\mathcal{P}_0(\Delta) + \mathcal{P}_1(\Delta) + \mathcal{P}_\perp(\Delta) = I \otimes I_{\mathbf{C}^2}, \quad I = \int_{-\infty}^{\infty} |\tau\rangle\langle\tau| d\tau, \quad (78)$$

$$\mathcal{P}_{0,1}(\Delta) = \left(\frac{1}{\sqrt{2}} \int_{-\Delta\tau}^{\Delta\tau + \tau_0} [f(\tau) + f(\tau - \tau_0)]|\tau\rangle d\tau\right) \times \left(\frac{1}{\sqrt{2}} \int_{-\Delta\tau}^{\Delta\tau + \tau_0} [f(\tau') + f(\tau' - \tau_0)]\langle\tau'| d\tau'\right) \otimes |e_{0,1}\rangle\langle e_{0,1}|, \quad (79)$$

and

$$\mathcal{P}_\perp(\Delta) = I - \mathcal{P}_0(\Delta) - \mathcal{P}_1(\Delta). \quad (80)$$

For the true extended states, the measurement (78)–(80) yields the results with the probabilities

$$\mathrm{Tr}\{\rho_{0,1}\mathcal{P}_{0,1}(\Delta)\} = 1 - O(e^{-\xi}), \quad \mathrm{Tr}\{\rho_{0,1}\mathcal{P}_\perp(\Delta)\} = O(e^{-\xi}). \quad (81)$$

By analogy with (67), (68) the measurement (78)–(80) on any states ρ not falling simultaneously into both intervals $(-\Delta\tau, \Delta\tau)$ and $(-\Delta\tau + \tau_0, \Delta\tau + \tau_0)$ yields

$$\int_{-\Delta\tau}^{\Delta\tau}\int_{-\Delta\tau}^{\Delta\tau} \delta_+(\tau - \tau')\rho(\tau, \tau')d\tau d\tau' = \frac{1}{2} - \frac{1}{2}e^{-\xi}, \quad \int_{-\Delta\tau + \tau_0}^{\Delta\tau + \tau_0}\int_{-\Delta\tau + \tau_0}^{\Delta\tau + \tau_0} \delta_+(\tau - \tau')\rho(\tau, \tau')d\tau d\tau' = \frac{1}{2} - \frac{1}{2}e^{-\xi}. \quad (82)$$

Thus, a time delay between states exceeding $2\Delta\tau$ leads to the probability of outcome for these states in the $\mathcal{P}_{0,1}$ channel dropping from almost unity (81) to nearly 1/2 (82):

$$\mathrm{Tr}\{\rho\mathcal{P}_{0,1}\} = \frac{1}{2} - \frac{1}{2}e^{-\xi}, \quad (83)$$

accordingly, the probability in the $\mathcal{P}_\perp$ channel increases from almost zero (80) to nearly 1/2 (82):

$$\mathrm{Tr}\{\rho\mathcal{P}_\perp\} = \frac{1}{2} - \frac{1}{2}e^{-\xi}. \quad (84)$$

Alice prepares (as in the previous case) Nk states and submits these states to the communication channels. As long as only half of the states is accessible to Bob, the probability for him to obtain information about the secret parity bit chosen by Alice does not exceed $(\Delta\tau \le \tau \le \Delta\tau + \tau_0)$

$$P_c(\text{parity}) \approx \frac{1}{2} + \left(\frac{1}{2} - O(e^{-\xi})\right)^{\alpha(N,k)Nk}. \quad (85)$$

For Alice, the probability of being undisclosed upon delay with the choice in at least one k-block does not exceed

$$\left(\frac{1}{2} - O(e^{-\xi})\right)^k. \quad (86)$$

The probability of successfully accomplishing the protocol (with the result obtained for all Nk states in the $\mathcal{P}_{0,1}$ channels) is

$$(1 - O(e^{-\xi}))^{\alpha(N,k)Nk} \quad (87)$$

that can be made arbitrarily close unity by properly selecting N, k, and ξ.

6. CONCLUSION

The existence of a limiting velocity of propagation for the quantum states allows the relativistic quantum protocols realizing the bit commitment and coin tossing scenarios to be formulated, in which the basic idea of such protocols is implemented, whereby one party presents another only a part of information (a part of the quantum state transmitted) about a secret bit chosen. However, a statistical nature of the results of measurements on the quantum states does not allow a fair-play protocol to be realized at a probability of exactly unity (at least, for the proposed scheme). Nevertheless, the fair-play protocol can be approached at a probability arbitrarily close to unity. In addition, the fundamentally nonlocal character of the quantum field states also poses limitations on the probability of trusted protocol outcomes within a finite time interval. However, the existence of arbitrarily strongly localized states admitted by the field theory basically allows a fair-play protocol to be constructed for an arbitrary time τ_0 (the time of secret bit keeping) with a preset probability arbitrarily close to unity.

In contrast to the nonrelativistic protocols, where only the structure of states in the Hilbert space is significant, the relativistic schemes involve explicit stages of the state preparation and propagation in the space–time between spatially separated parties. Since the spin and helicity states in the nature cannot exist outside the spatial degrees of freedom of a quantum system, an allowance for these degrees of freedom increases possibilities for the construction of quantum cryptographic protocols.

It should be emphasized that the proposed protocol employs orthogonal states. The probability of an error in distinguishing orthogonal states is related to the fact that a measurement over such states may give no result at all (a classical instrument yields no response, e.g., the pointer does not deviate) if the transmitted state falls outside the space–time region in which the instrument is located. Therefore, there are three possible outcomes of the measurements: the classical instrument either operates in one of the two channels ($\mathcal{P}_0$ or $\mathcal{P}_1$) or does not operate at all. Once the instrument has operated, the states are unambiguously distinguished. If only a part of the state is accessible, there is a nonzero probability that the instrument will not operate at all; the smaller the accessible part of the state, the closer this probability to unity. In this case, the receiving party may only simply guess the state (accordingly, the probability of an error in determining the outcome of such a measurement performed with a classical instrument is 1/2).

Since the spin and helicity states cannot exist outside the spatial degrees of freedom, a limited access to the coordinate space automatically restricts the access to the Hilbert space of states. A situation is even possible when the state of a system would become inaccessible at all (the state amplitude will be zero in the spatial region accessible for the measurements).

It should be noted that this situation is different from that considered by Tal Mor [41] in the context of discussion of the Goldenberg–Vaidman quantum cryptography scheme based on orthogonal states [18]. For a pair of orthogonal states of a composite system comprising two subsystems *a* and *b* with the space of states $\mathcal{H}_a \otimes \mathcal{H}_b$, we have

$$|\psi_0\rangle = \alpha_0|\phi_0(a)\rangle \otimes |\phi_0(b)\rangle + \beta_0|\phi_1(a)\rangle \otimes |\phi_1(b)\rangle,$$

$$|\psi_1\rangle = \alpha_1|\phi_0(a)\rangle \otimes |\phi_0(b)\rangle + \beta_1|\phi_1(a)\rangle \otimes |\phi_1(b)\rangle,$$

where the coefficients $\alpha_{0,1}$ and $\beta_{0,1}$ are such that the states $|\psi_0\rangle$ and $|\psi_1\rangle$ are orthogonal:

$$\langle\psi_0|\psi_1\rangle = 0.$$

If only one subsystem (e.g., $\mathcal{H}_a$) is accessible, the states of the other subsystem *a* are nonorthogonal,

$$\rho_1 = \mathrm{Tr}_{\mathcal{H}_a}\{|\psi_1\rangle\langle\psi_1|\}, \quad \rho_0 = \mathrm{Tr}_{\mathcal{H}_a}\{|\psi_0\rangle\langle\psi_0|\},$$
$$\mathrm{Tr}_{\mathcal{H}_b}\{\rho_0\rho_1\} \neq 0,$$

and, hence, undistinguishable. In our case, the states remain orthogonal even when bounded in a subspace and may be undistinguishable only due to some features of the space–time structure.

The proposed protocol can be generalized to a noisy channel [42], since the initial orthogonality of states allows the classical codes to be employed [43].

In the given scheme, the protocol duration ($\approx\tau_0$) is determined by the effective extension of states, which can be evaluated for photons using the width of the frequency spectrum. At present, a minimum width of the spectrum in the optical range, attained in circular fiber cavities, amounts to $\Delta\omega \approx 10$ kHz. The corresponding effective state length is

$$L \approx c/\Delta\omega = 3 \times 10^{10}/10^4 \text{ cm} = 3 \times 10^6 \text{ cm (30 km)}.$$

Accordingly, the protocol duration is

$$\tau_0 \approx 1/\Delta\omega \approx 10^{-3} \text{ s}.$$

Although there are no fundamental constraints that would forbid the time τ_0 to be made arbitrarily large (and, accordingly, $\Delta\omega$ arbitrarily small), this task is technically very difficult. However, this circumstance is insignificant for the coin tossing protocol were the time τ_0 for obtaining a trusted lot can be arbitrary. In contrast, the τ_0 value is significant in the bit commitment protocol where this parameter determines the time during which the secret bit can be kept. This situation is quite generally encountered in the experimental realization of various systems for the transfer and processing of "quantum information," when certain basic possibilities, allowed by the laws of quantum mechanics, are very difficult to realize using the existing technical facilities.

It should be noted that real data transmission over long distances is performed using fiber optic systems in which the signals propagate at a velocity somewhat lower compared to the speed of light in vacuum. This circumstance is by no means restrictive, the only requirement being that the time separation of state "halves" would exceed the channel length divided by the light velocity in a given optical fiber.

ACKNOWLEDGMENTS

The work was supported by the Russian Foundation for Basic Research (project no. 99-02-18127), by the Project "Physical Principles of Quantum Computer," and by the Federal Program "Advanced Technologies and Devices of Micro- and Nanoelectronics" (project no. 02.04.5.2.40.T.50).

REFERENCES

1. L. Landau and R. Peierls, Z. Phys. **69**, 56 (1931); L. D. Landau, *A Collection of Scientific Works* (Nauka, Moscow, 1969), Vol. 1.

2. N. Bohr and L. Rosenfeld, Math.-Fys. Medd. **12**, 3 (1933); N. Bohr, *A Collection of Scientific Works* (Nauka, Moscow, 1971), Vol. 1.
3. V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics* (Nauka, Moscow, 1982).
4. W. K. Wootters and W. H. Zurek, Nature **299**, 802 (1982).
5. C. H. Bennett, Phys. Rev. Lett. **68**, 3121 (1992); C. H. Bennett, G. Brassard, and N. D. Mermin, Phys. Rev. Lett. **68**, 557 (1992).
6. S. Wiesner, SIGACT News **15**, 78 (1983).
7. C. H. Bennett and G. Brassard, in *Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing* (IEEE, New York, 1984), p. 175.
8. A. Ekert, Phys. Rev. Lett. **67**, 661 (1991).
9. G. Brassard, C. Crépeau, R. Jozsa, and D. Langlois, in *Proceedings of the 34th Annual IEEE Symposium on the Foundation of Computer Science* (IEEE Comput. Soc. Press, Los Alamitos, 1993), p. 362.
10. G. Brassard and C. Crépeau, in *Advances in Cryptology: Proceedings of Crypto'90* (Springer-Verlag, Berlin, 1991), Lecture Notes in Computer Science, Vol. 537, p. 49.
11. M. Ardehali, quant-ph/9603015.
12. D. Mayers, L. Salvail, and Y. Chiba-Kohno, quant-ph/9904078.
13. M. Blum, in *Proceedings of the 24th IEEE Computer Conference, 1982*, p. 133; SIGACT News **15**, 23 (1983).
14. P. W. Shor, in *Proceedings of the 35th Annual IEEE Symposium on Foundations of Computer Science, Santa Fe, 1994*, Ed. by S. Goldwasser (IEEE Comput. Soc. Press, Los Alamitos, 1994), p. 124.
15. A. Yu. Kitaev, Usp. Mat. Nauk **52**, 54 (1997).
16. H.-K. Lo and H. F. Chau, Phys. Rev. Lett. **78**, 3410 (1997).
17. D. Mayers, Phys. Rev. Lett. **78**, 3414 (1997).
18. L. Goldenberg and L. Vaidman, Phys. Rev. Lett. **75**, 1239 (1995); quant-ph/9506030.
19. A. Peres, quant-ph/9509003.
20. M. Koashi and N. Imoto, Phys. Rev. Lett. **79**, 2383 (1997).
21. S. N. Molotkov and S. S. Nazin, quant-ph/0008008.
22. R. Laiho, S. N. Molotkov, and S. S. Nazin, quant-ph/0006010.
23. R. Laiho, S. N. Molotkov, and S. S. Nazin, quant-ph/0005067; quant-ph/0005068; Phys. Lett. A **275**, 37 (2000).
24. A. Kent, quant-ph/9810067; quant-ph/9810068; quant-ph/9906103; Phys. Rev. Lett. **83**, 1447 (1999).
25. S. N. Molotkov and S. S. Nazin, quant-ph/9911055; quant-ph/9910034; Zh. Éksp. Teor. Fiz. **117**, 818 (2000) [JETP **90**, 714 (2000)]; Pis'ma Zh. Éksp. Teor. Fiz. **70**, 684 (1999) [JETP Lett. **70**, 702 (1999)].
26. Yu. A. Brychkov and A. P. Prudnikov, *Integral Transforms of Generalized Functions* (Nauka, Moscow, 1977; Gordon and Breach, New York, 1989).
27. N. N. Bogolyubov, A. A. Logunov, A. I. Oksak, and I. T. Todorov, *General Principles of Quantum Field Theory* (Nauka, Moscow, 1987).
28. N. N. Meĭman, Zh. Éksp. Teor. Fiz. **47**, 1966 (1964) [Sov. Phys. JETP **20**, 1320 (1964)].
29. D. A. Kirzhnits, Usp. Fiz. Nauk **90**, 129 (1966) [Sov. Phys. Usp. **9**, 692 (1966)].
30. A. M. Jaffe, Phys. Rev. **158**, 1454 (1967).
31. N. N. Bogolyubov, A. A. Logunov, and I. T. Todorov, *Introduction to Axiomatic Quantum Field Theory* (Nauka, Moscow, 1969; Benjamin, New York, 1975).
32. I. Bialynicki-Birula, Phys. Rev. Lett. **80**, 5247 (1998).
33. N. Wiener and R. Paley, *Fourier Transform in the Complex Domain* (American Mathematical Society, New York, 1934; Nauka, Moscow, 1964).
34. C. W. Helstrom, Inf. Control. **10**, 254 (1967); C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic, New York, 1976; Mir, Moscow, 1979), Mathematics in Science and Engineering, Vol. 123.
35. A. P. Prudnikov, Yu. A. Brychkov, and O. I. Marichev, *Integrals and Series* (Nauka, Moscow, 1981; Gordon and Breach, New York, 1986).
36. C. A. Fuchs, quant-ph/9601020.
37. C. H. Bennett, Tal Mor, and J. Smolin, Phys. Rev. A **54**, 2675 (1996); Tal Mor, quant-ph/9906073.
38. C. E. Shannon, Bell Syst. Tech. J. **27**, 397, 623 (1948).
39. I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems* (Akademiai Kiado, Budapest, 1981).
40. S. N. Molotkov and S. S. Nazin, Pis'ma Zh. Éksp. Teor. Fiz. **63**, 646 (1996) [JETP Lett. **63**, 687 (1996)]; A. V. Krasheninnikov, S. N. Molotkov, S. S. Nazin, and L. A. Openov, Zh. Éksp. Teor. Fiz. **110**, 1257 (1997) [JETP **85**, 682 (1997)]; K. M. Gheri, C. Saavedra, P. Törmä, *et al.*, Phys. Rev. A **58**, R2627 (1998).
41. Tal Mor, Phys. Rev. Lett. **80**, 3137 (1998).
42. S. N. Molotkov and S. S. Nazin, Pis'ma Zh. Éksp. Teor. Fiz. **73**, 114 (2001) [JETP Lett. **73**, 107 (2001)].
43. E. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes* (North-Holland, Amsterdam, 1977).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 93, No. 4, 2001, pp. 895–902.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 4, 2001, pp. 1027–1036.

MISCELLANEOUS

On the Applicability of the Frenkel–Kontorova Model to Describing the Dynamics of Vacancies in a Polymeric Crystal Chain

E. A. Zubova

Semenov Institute of Chemical Physics, Russian Academy of Sciences, ul. Kosygina 4, Moscow, 117977 Russia
e-mail: zubova@center.chph.ras.ru

Received April 10, 2001

Abstract—The necessary conditions of the applicability of the Frenkel–Kontorova one-dimensional model [the approximation of immobile neighboring chains plus sine-Gordon (continual) equation for nonlinear dynamics of the chain under consideration] to describing the dynamics of vacancies in a polymeric crystal chain are determined. It is shown that these conditions are satisfied for polyethylene crystals. The physical mechanism of model applicability limitations is established. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION: ONE-DIMENSIONAL MODELS FOR DESCRIBING NONLINEAR DYNAMICS IN THREE-DIMENSIONAL SYSTEMS

One of the most important achievements of nonlinear physics is the possibility of describing localized waves that propagate at a constant velocity, that is, solitons or soliton-like excitations. Such waves are often observed in physical systems, but they are seldom truly one-dimensional. These are either waves in filamentary systems (a magnetic flux quantum in a long and narrow Josephson contact, a solitary surface wave in a shallow and narrow channel with water, or an optical pulse of a picosecond width in a thin optical fiber waveguide made of a nonlinear material) or plane waves (a turning wave in a uniaxial ferromagnet with anisotropy of the type of an easy magnetization plane, ion–sound waves in a homogeneous collisionless nonisothermal plasma and magnetoacoustic waves in a cold plasma placed into a magnetic field, or an optical pulse in a nonlinear medium). In this context, examples of quasi-one-dimensional waves virtually localized along a line and propagating in a three-dimensional (3D) medium like truly one-dimensional excitations are very interesting. For instance, a soliton of stretches in a polymeric crystal (that is, a chain unit vacancy without rupture of internal bonds localized in a small chain portion) is a quasi-one-dimensional soliton.

Solutions corresponding to nontopological solitons in multidimensional systems such as Langmuir (electron) waves in a cold plasma are often unstable with respect to the wave collapse, although sometimes, new stable solitons localized in all directions appear, for instance, as magnetoacoustic waves in a cold 2D plasma. The problem of taking into account the 3D character of real physical objects for topological solitons, which are solutions to equations of the sine-Gordon type, is qualitatively different in nature. The turning region in a one-dimensional ferromagnet with anisotropy of the type of an easy magnetization plane, an edge dislocation in a low-molecular-weight crystal, or a chain unit vacancy in a polymeric crystal chain should exist as static objects (structural defects) in the 3D as well as in one-dimensional case. In a 3D system, only the type of their dynamic behavior may change from soliton-like (motion at a constant velocity in a cold crystal) to pinning because of a lowering of the upper bound of the spectrum of velocities.

In this work, we study the conditions of the applicability of the one-dimensional Frenkel–Kontorova model to analyzing nonlinear dynamics of a topological localized soliton-like excitation in a 3D system for the example of a vacancy in a polymeric crystal chain. We also consider the physical mechanisms responsible for changes in dynamic behavior of structural defects described above.

2. POINT STRUCTURAL DEFECTS IN POLYMERIC CRYSTAL CHAINS

The energy characteristics and the type of the dynamics of structural defects determine the relaxation properties and the special features of phase transitions in crystals. For this reason and in view of the availability of polymers with high crystallinity degrees, many studies of point and line structural defects in crystals formed by chain molecules have appeared during recent years.

Because of strong anisotropy and hierarchy of interactions (intrachain covalent chemical bonds are several orders of magnitude more rigid than interchain van der Waals bonds), vacancies with rupture of intrachain

1063-7761/01/9304-0895$21.00

covalent bonds are virtually immobile. Such crystals can, however, contain other, specially polymeric, point defects caused by chain deformations rather that the rupture of intrachain bonds. These defects may be localized on a small chain portion, for instance, a chain can contain a chain unit vacancy or (a plane zigzag chain) a point defect of rotation through 180° accompanied by chain elongation or contraction by half the chain period to preserve crystallographic order outside the defect region.

The idea that polymeric crystal chains can contain torsional defects with elongation was originally advanced in [1, 2] in relation to dielectric α-relaxation in weakly oxidized polyethylene; the exact form of this defect was established in [3]. The model including Brownian movement of such a defect along the chain as a relaxation mechanism (see review [4] and two series of works, [2, 5] and [6]) allowed some special features of the process to be explained, namely, its occurrence in the crystalline fraction, anisotropy with respect to the direction of the applied field vector, the local character of the mechanism, the presence of the α-peak in dielectric relaxation of polyethylene and isotactic polypropylene and its absence for syndiotactic polypropylene and isotactic polystyrene [7]. Since then, the dynamics of structural defects of this type in polymeric crystals have been studied fairly extensively [3, 6–10].

Movement of a point structural defect along a chain in a polymeric crystal is usually treated in the quasi-one-dimensional approximation of immobile neighboring chains and is described in terms of topological soliton-like excitations in the continual approximation as a nonlinear wave which freely propagates at a constant subsonic velocity along the chain and changes the state of the chain after its passage.

Calculations of the characteristics of solitons even in the approximation of immobile neighboring chains requires knowledge of the form of the interatomic interaction potential and its parameters, which cannot be obtained from experimental data, but which to a substantial degree determine the type of behavior of defects. The idea that expanding the interchain potential in a polymeric crystal into the Fourier series can be performed analytically by considering the potential of a linear chain of atoms and summing their contributions by the Poisson summation rule was long ago advanced by McCullough [11]. This approach was used to estimate the temperature dependence of shear moduli in paraffins [12] and to calculate crystal potential energy minima corresponding to the monoclinic and orthorhombic polyethylene polymorphs [13]. The potential of a linear chain of atoms was calculated not only for Lennard-Jones-type interatomic interactions but also for slowly decreasing Coulomb interactions (see review [14]). More recently [15], this method was used to determine the form of the interatomic potential in polyethylene.

In all these works, the periodic potential of a linear chain of atoms was truncated after the first harmonic of the obtained Fourier expansion, $A_0 + A_1\cos(2\pi z/c)$, where c is the distance between chain atoms. The correctness of this approximation was checked by trial calculations of the amplitude of the next harmonic. Physical criteria of the applicability of such an approximation (the sine-Gordon equation for describing nonlinear chain dynamics) to various real crystals were not, however, formulated.

On the other hand, it is not quite clear whether or not the environment of a chain can be considered immobile in analyzing the dynamics of chain point defects. Molecular dynamics simulations show [16, 17] that mobility of neighboring chains can have a noticeable effect on the dynamics of defects. In several works (e.g., see [18]), mobility of neighboring chains was taken into account by phenomenologically introduced terms which, in the limit of stationary neighboring chains, corresponded to the approximation of the first harmonic in the Fourier expansion, but the physical meaning of these terms was not discussed.

In this work, we make an attempt to fill up the two gaps specified above and determine the limits of the applicability of the Frenkel–Kontorova one-dimensional model (the approximation of immobile neighboring chains plus the sine-Gordon equation for the nonlinear dynamics of the chain under consideration) to describing the dynamics of vacancies in a polymeric crystal chain (Sections 4 and 3, respectively). An example of a real polymeric crystal which satisfies these criteria and to which the Frenkel–Kontorova one-dimensional model is applicable is described in Section 5.

3. A PERIODIC POTENTIAL OF A LINEAR CHAIN OF ATOMS

Consider potential $W(c, b, z)$ created on axis z by a linear chain of rigidly fixed atoms situated parallel to this axis at distance b; c is the interatomic distance (Fig. 1). The analysis will be performed specifically for van der Waals interactions between a "test" atom and every other chain atom,

$$U_{LJ}(r) = U_0 u(r) = U_0\left(\frac{r_0}{r}\right)^6\left[\left(\frac{r_0}{r}\right)^6 - 2\right].$$

Here, U_0 is the potential energy minimum and r_0 is the point at which this minimum is attained. The summation over all atoms yields

$$W(c, b, z) = \sum_{l=-\infty}^{\infty} U_{LJ}\left(\sqrt{b^2 + \left(\frac{c}{2} + lc - z\right)^2}\right). \qquad (1)$$

If the equilibrium position of the test atom at the $z = 0$ point is largely determined by interactions with its two

nearest neighbors (atoms with numbers 0 and –1), the "local" case, then

$$\varepsilon_U \equiv \left|\frac{U_{LJ}(l=1)}{U_{LJ}(l=0)}\right| \ll 1.$$

Let us introduce the $y = b/r_0$ dimensionless variable and the $\delta = c/2r_0$ parameter and write the terms of (1) in an explicit form. In "local" equilibrium, when $y^2 + \delta^2 \approx 1$, the ε_U value only depends on parameter δ. Physically meaningful are δ values in the (0, 1) interval. At a constant c parameter, the optimal b distance from the neighboring chain tends to infinity when $\delta \longrightarrow 0$ and to zero when $\delta \longrightarrow 1$.

Selecting $\varepsilon_U = 0.03$ as a boundary of the local case, we find that the potential of the atom at $z = 0$ is in reality only determined by two nearest atoms from the neighboring chain if $\delta > \sqrt{3/8} \approx 0.61$.

Otherwise (when the contributions of the other atoms are large), sum (1) can be calculated by the Poisson summation formula,

$$\sum_{n=-\infty}^{\infty} F(nh) = \int_{-\infty}^{\infty} \frac{d\zeta}{h} F(\zeta) \sum_{m=-\infty}^{\infty} \exp\left(2i\pi m \frac{\zeta}{h}\right).$$

Then,

$$W(c, b, z) = \int_{-\infty}^{\infty} \frac{dx}{c} U_{LJ}(\sqrt{b^2 + x^2}) \times \left[1 - 2\cos\left(\frac{2\pi}{c}x\right)\cos\left(\frac{2\pi}{c}z\right) + 2\cos\left(\frac{2\pi}{c}2x\right)\cos\left(\frac{2\pi}{c}2z\right) - 2\cos\left(\frac{2\pi}{c}3x\right)\cos\left(\frac{2\pi}{c}3z\right) + \ldots\right]. \quad (2)$$

All these integrals are easily found by the theory of residues, and the expansion of the $W(\delta, y, z)$ potential periodic along z into a Fourier series takes the form

$$W(\delta, y, z) = \frac{3\pi}{8\delta} U_0 \left[W_0(y) + f(\delta, y)\cos\left(\frac{2\pi}{c}z\right) - f\left(\frac{\delta}{2}, y\right)\cos\left(\frac{2\pi}{c}2z\right) + f\left(\frac{\delta}{3}, y\right)\cos\left(\frac{2\pi}{c}3z\right) - \ldots\right]. \quad (3)$$

Here, the main contribution to the potential only depends on the interchain distance and is independent of either z or δ,

$$W_0(y) = \frac{1}{y^5}\left(\frac{21}{64}\frac{1}{y^6} - 1\right).$$

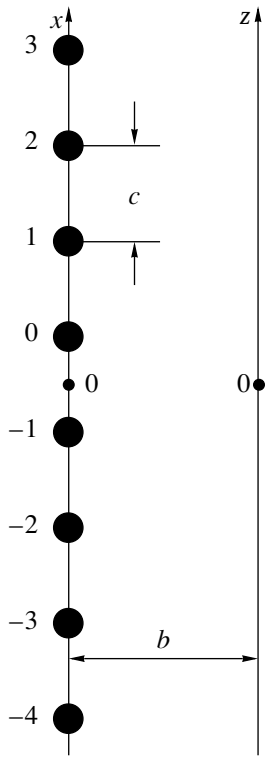

Fig. 1. Linear chain of atoms with fixed positions along the x axis. Each atom interacts with the test atom situated on the z axis according to the Lennard-Jones potential. The total potential energy of the test atom is found as a function of chain period c, distance between axes b, and test atom position z.

The amplitudes of the harmonics are given by the $f(\delta, y)$ function,

$$f(\delta, y) = -\frac{2}{3}\exp\left(-\pi\frac{y}{\delta}\right)\frac{1}{y^5}\left(\frac{Q_5(\pi y/\delta)}{8y^6 \times 5!} - Q_2\left(\pi\frac{y}{\delta}\right)\right),$$

$$Q_5(p) = p^5 + 15p^4 + 105p^3 + 420p^2 + 945p + 945,$$

$$Q_2(p) = p^2 + 3p + 3.$$

The condition of the applicability of expansion (3) is the smallness of the ratio of the first harmonic amplitude to the main term. This condition determines the boundary of the "collective" case: the requirement $|f(\delta, y_0)/W_0(y_0)| < 0.1$, where y_0 is the position of the $W_0(y)$ function minimum, is equivalent to the condition $\delta < 0.41$. The amplitude of every next harmonic is then much less than that of the preceding harmonic because of the exponential dependence of the f function on the harmonic number. Even at $\delta = 0.41$, the amplitude of the second harmonic is 100 times lower than that of the first one. If we retain the first harmonic in the expansion (this is necessary to obtain the dependence of the chain potential on z), the weaker condition $|f(\delta/2, y_0)/f(\delta, y_0)| < 0.1$ extends the boundary of the collective case to $\delta = 0.55$.

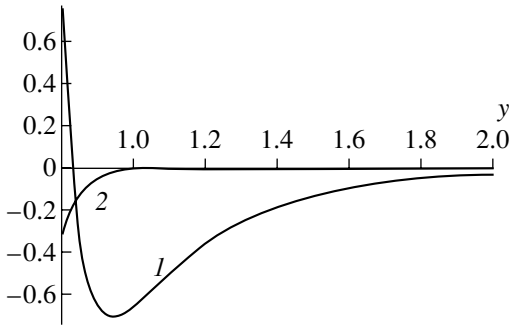

Fig. 2. Main contribution W_0 (curve *1*) to the potential of a linear chain of atoms (Fig. 1) and the $f(\delta = 0.3, y)$ amplitude (curve *2*) of the first harmonic along z as functions of dimensionless distance $y = b/r_0$ (r_0 is the equilibrium Lennard-Jones interatomic potential distance) between axes x and z.

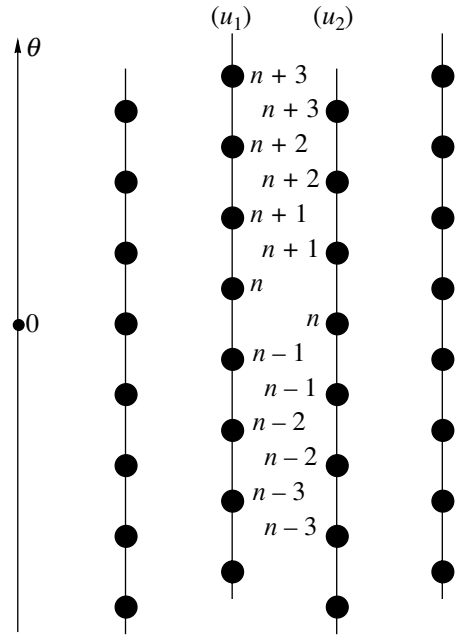

Fig. 3. Equilibrium two-dimensional crystal formed by linear chains.

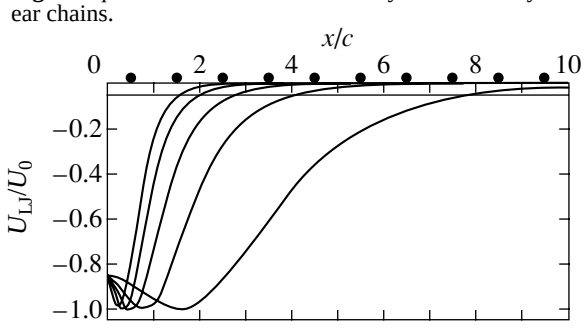

Fig. 4. Potential created by one chain atom situated at point x (Fig. 1) at the position of the test atom situated at $z = 0$ on the z axis as a function of dimensionless distance x/c (c is the chain period). At the selected $b = r_0 y_0$ interchain distance, the main contribution of the whole chain to the potential is minimum. Curves for different δ parameter values are plotted, from left to right: $\delta = 0.5$, 0.4, 0.3, 0.2, and 0.1. Solid circles along the x/c axis are real atomic positions on the axis. The horizontal line corresponds to the –0.05 level.

We see that, in the collective case, the z dependence of the potential generated on the axis of a linear chain of atoms is largely determined by one harmonic. This distinguishes the collective case from the local case, in which the periodic substrate is created by flat wells with narrow barriers in between. It follows that the boundary of the collective case is also the boundary of the applicability of the sine-Gordon equation to describing the nonlinear dynamics of a chain in a polymeric crystal.

Understanding the type of the polymeric crystal under study is necessary for correctly estimating the type of mobility of point defects in its chains. It is known [19] that, if barriers are narrow, a much larger interchain rigidity is required for kinks to move along chains without pinning than when there is a single harmonic.

The $W_0(y)$ and $f(\delta = 0.3, y)$ functions are plotted in Fig. 2. The $W_0(y)$ function attains a minimum at

$$y = y_0 = \left(\frac{21}{64} \times \frac{11}{5}\right)^{1/6} \approx 0.94713161\ldots .$$

The $f(\delta = 0.3, y)$ function is negative at reasonable y values (it takes on exceedingly small positive values of the order of 10^{-6} at $y > 1.4$). This means that, if we wish to construct a two-dimensional crystal of linear chains, we must displace atoms of every next chain by half the period with respect to atoms of the preceding chain (Fig. 3).

Let us estimate the number of particles that make a noticeable (for instance, larger than 5% of U_0) contribution to the potential in the collective case. At $z = 0$ and $b = y_0 r_0$, the potential created by a chain atom at point x (Fig. 1) is given by

$$\frac{U_{LJ}(\sqrt{(r_0 y_0)^2 + x^2})}{U_0} = \frac{1}{y_0^6}\left[1 + \left(\frac{x}{c}\frac{2\delta}{y_0}\right)^2\right]^{-3}$$

$$\times \left\{\left[1 + \left(\frac{x}{c}\frac{2\delta}{y_0}\right)^2\right]^{-3}\frac{1}{y_0^6} - 2\right\}.$$

The plots of this function obtained at various δ values are given in Fig. 4. As expected, four particles make noticeable contribution to the potential ($N_s = 4$) if $\delta \sim 0.4$–0.5. The data on the other δ values are given in the table.

Let us estimate the ratio between the amplitude of the first harmonic in (3) to the lowest Lennard-Jones interatomic potential energy at various δ. These ratios, $(3\pi/8\delta)|f(\delta, y_0)|$, are also given in the table. We see that, even at the boundary ($\delta = 0.5$), the amplitude of the substrate is smaller than $0.25U_0$, whereas at arbitrary $\delta > 0.61$ (the local case), half the difference of potential hump and well energies exceeds $4.7U_0$.

This is an important difference between the local and collective cases. The width of a topological kink is

Dependence of the N_s number of atoms actually forming potential at the $z = 0$ point (Fig. 1) and the amplitude of the first harmonic of this potential on δ

δ	0.1	0.2	0.3	0.4	0.5
N_s	8×2	4×2	3×2	2×2	2×2
$\frac{3\pi}{8\delta}\lvert f(\delta, y_0)\rvert$	7×10^{-8}	4.45×10^{-3}	7.33×10^{-2}	0.191	0.239

proportional to the square root of the ratio between intrachain and substrate rigidities. At a large substrate amplitude (even if the substrate is sinusoidal), substantial intrachain rigidity is required for kinks to remain fairly broad and experience no deceleration caused by emission of energy into chain phonon modes [20].

It follows that the transition from the collective to the local case causes an increase in the amplitude of the linear chain potential and changes its form from purely sinusoidal to broad wells separated by narrow barriers. This corresponds to the transition from the soliton type of mobility of point defects to emission of phonons by defects (and the corresponding lowering of the upper spectrum bound) and, eventually, to purely diffusive mobility type with pinning.

4. TAKING INTO ACCOUNT MOBILITY OF NEIGHBORING CHAINS. INTERCHAIN INTERACTION POTENTIAL

The interaction energy between two chains whose atoms are displaced by $(u_1)_n$ and $(u_2)_n$ from their equilibrium positions shown in Fig. 3 is usually written in the form [18]

$$W(u_1, u_2) = \varepsilon_0 \int_{-\infty}^{\infty} \frac{du}{c} \times \left\{ -\alpha\left[1 - \cos\left(\frac{2\pi}{c}(u_1 - u_2)\right)\right] + \beta \frac{\partial u_1}{\partial y}\frac{\partial u_2}{\partial y} \right\}. \quad (4)$$

Setting u_2 equal to zero, we obtain the sine-Gordon equation for the nonlinear dynamics of the first chain, u_1. The criterion of the applicability of this equation was discussed in Section 3. Let us add interaction energy (4) to the Hamiltonian

$$H_0 = \varepsilon_0 \int_{-\infty}^{\infty} \frac{dy}{c} \left\{ \frac{1}{2}\left(\frac{\partial u_1}{\partial t}\right)^2 + \frac{1}{2}\left(\frac{\partial u_1}{\partial y}\right)^2 + \left[1 - \cos\left(\frac{2\pi}{c}u_1\right)\right] + \frac{1}{2}\left(\frac{\partial u_2}{\partial t}\right)^2 + \frac{1}{2}\left(\frac{\partial u_2}{\partial y}\right)^2 + \left[1 - \cos\left(\frac{2\pi}{c}u_2\right)\right] \right\} \quad (5)$$

and consider the dynamics of a kink in chain u_1. Assuming that $\alpha \ll 1$ and $\beta \ll 1$, one can find [18] within the framework of the perturbation theory that interaction (4) between chain u_1 and mobile chain u_2 changes the shape of the kink and causes the appearance of a perturbation in chain u_2 which moves together with the kink. The amplitude of this perturbation is proportional to $\alpha + \beta/(1 - v^2)$, where v is the velocity of the kink in units of v_s, the sound velocity of an isolated chain. The amplitude increases infinitely as the velocity of the kink approaches the v_s value, which shows that such perturbation theory cannot be used to take into account kink radiation loss which appears when its velocity increases. Indeed, it is easy to show that, in a system of two interacting chains, the introduction of the $\varepsilon_0\beta(\partial u_1/\partial y)(\partial u_2/\partial y)$ term results in that the lowest phase velocity of phonons common to both chains equals $v_{1,2} = \sqrt{1 \pm \beta}$ (in v_s units) for two branches corresponding to phonons synphase and antiphase in chains, respectively. It follows that, regardless of the sign of β, the velocity of phonons on one branch is always smaller than unity, and if a kink exceeds this velocity, Vavilov–Cherenkov-type radiation appears. Conversely, the term

$$-\varepsilon_0\alpha\left[1 - \cos\left(\frac{2\pi}{c}(u_1 - u_2)\right)\right]$$

only influences the width of the phonon spectrum gap rather than v_1 and v_2. We therefore see that two terms in (4) have different physical meanings.

Note also that Eq. (4) is internally contradictory. Indeed, the expression for energy only contains one harmonic, and this expression is obtained in the continual approximation, in which displacements $(u_1)_n$ and $(u_2)_n$ are replaced by fields $u_1(y)$ and $u_2(y)$, and the summation over atoms is replaced by the integration over the chain. This is the result of using the Poisson summation formula. The condition that allows us to retain a single harmonic in the obtained Fourier expansion of the potential is the collective character of interactions, when substrate for a given atom is formed by many neighboring chain atoms. On the other hand, it follows from (4) that the interaction between the fields is quasilocal; that is, the equation for field $u_1(y_0)$ contains field u_2 and its second derivative with respect to the coordinate only at the same y_0 point.

As (4) is extensively used to study interactions between kinks in various polymeric crystal chains and the influence of the mobility of neighboring chains on the dynamics of kinks [18] and to describe the dynamics of a fluxon in one of two interacting Josephson lines (see [21–23] and the references therein), it is interesting to determine if there exists a region of system parameters in which this equation is not physically meaningless.

The interaction energy between two chains in a crystal (see Fig. 3) is the sum of interactions of all pairs of atoms,

$$W_{1-2} = \sum_{n=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} U_{LJ} \times\left(\sqrt{b^2+\left[\frac{c}{2}+lc+(u_1)_{n+l}-(u_2)_n\right]^2}\right).$$

Replacing the summation over l by integration over θ in the Poisson formula,

$$W_{1-2} \approx U_0\sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{d\theta}{c} \times u\left(\sqrt{b^2+(\theta+u_1(nc+\theta)-u_2(\theta))^2}\right) \times\left[1-2\cos\left(\frac{2\pi}{c}\theta\right)+2\cos\left(\frac{2\pi}{c}2\theta\right)-\ldots\right], \quad (6)$$

we see that terms depending on the difference of fields $u_1(nc) - u_2(nc)$ at a single point nc can only appear in the expression for energy if, in calculating the integral over θ, $u_1(nc + \theta)$ can be expanded into a Taylor series and the corrections depending on θ can be assumed to be small,

$$u_1(nc+\theta) \approx u_1(nc) + \theta u_1'(nc) + \frac{\theta^2}{2}u_1''(nc) + \ldots, \quad (7)$$

that is, on the segment of length $N_s c/2$, where the $u(\sqrt{b^2+\theta^2})$ function strongly varies (Fig. 4), the $u_1(\theta)$ function should be almost constant. If we are interested in the dynamics of vacancies, the characteristic spatial scale of substantial changes in $u_1(\theta)$ is the $L_k c$ kink half-width. The condition formulated above is then written as

$$L_k \gg N_s/2. \quad (8)$$

If this condition is satisfied, we can change variables

$$\zeta \equiv u_1(nc) - u_2(nc) + \theta(1+u_1'(nc)) + \frac{\theta^2}{2}u_1''(nc),$$

$$d\zeta = d\theta(1+u_1'(nc)+\theta u_1''(nc)),$$

$$\theta \approx \frac{\zeta-[u_1(nc)-u_2(nc)]}{1+u_1'(nc)} \approx \zeta-[u_1(nc)-u_2(nc)],$$

$$d\theta \approx d\zeta\{1-u_1'(nc) - [\zeta-(u_1(nc)-u_2(nc))]u_1''(nc)+\ldots\}$$

to obtain

$$W_{1-2} \approx U_0\sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{d\zeta}{c} \times\{1-u_1'(nc)+[u_1(nc)-u_2(nc)]u_1''(nc)\}u(\sqrt{b^2+\zeta^2}) \times\left[1-2\cos\left(\frac{2\pi}{c}\zeta\right)\cos\left(\frac{2\pi}{c}[u_1(nc)-u_2(nc)]\right)+\ldots\right].$$

Comparing this equation with (2) and (3), replacing the summation over n by the integration along the chain, and only retaining the largest terms describing the interaction of the fields, we obtain

$$W_{1-2} \approx U_0\frac{3\pi}{8\delta}\int_{-\infty}^{\infty}\frac{d\xi}{c} \times\left\{W_0u_1'(\xi)u_2'(\xi)+f\cos\left(\frac{2\pi}{c}[u_1(\xi)-u_2(\xi)]\right) - fu_1'(\xi)\cos\left(\frac{2\pi}{c}[u_1(\xi)-u_2(\xi)]\right)\right\}. \quad (9)$$

The last term in this equation is seemingly asymmetric with respect to fields u_1 and u_2, but it can easily be reduced to the symmetrical form

$$-\frac{1}{2}f[u_1'(\xi)+u_2'(\xi)]\cos\left(\frac{2\pi}{c}[u_1(\xi)-u_2(\xi)]\right)$$

by adding the total derivative with respect to the coordinate to energy density. The first term is transformed to the symmetrical form via the integration by parts. On the assumptions made above, the third term is much smaller than the first two ones.

The first two terms in Eq. (9) for the energy density are familiar to us because they are present in (4). However in (9), "phenomenological" constants are expressed in terms of interatomic potential parameters,

$$\varepsilon_0\alpha = U_0\frac{3\pi}{8\delta}f(\delta,y) < 0,$$

$$\varepsilon_0\beta = U_0\frac{3\pi}{8\delta}W_0(y) < 0,$$

$$\frac{\alpha}{\beta} \ll 1.$$

We see that, in the terminology of [18], the interaction between two linear chains of atoms that we are considering is interchain attraction. We now know which of many variants of interactions between kinks in neighboring chains considered in [18] (depending on signs of and the ratio between α and β) corresponds to the Lennard-Jones interatomic potential.

It is known that a discrete linear chain of atoms linked by springs and lying on a sine substrate (for instance, created by immobile neighboring chains) can be described by the continual sine-Gordon equation if the static kink half-width in interatomic distance units, L_k, which is proportional to the square root of the ratio between the rigidities of springs and substrate, is much larger than one. Numerical simulation [24] shows that, already at $L_k \approx 2$–4, the kink virtually does not emit and moves at an almost constant velocity (at a not too high velocity, $v_k = 0.5v_s$). However, if the substrate is created by mobile rather than immobile chains, it is necessary that the interaction energy between chains be representable in form (9) for the kink not to radiate; that is, the $\beta(\partial u_1/\partial y)(\partial u_2/\partial y)$ term resulting in radiative friction should be small. This requires condition (8) to be satisfied, namely, the kink half-width in the chain should greatly exceed not unity but half the number of particles, N_s, that really form substrate for each atom. The table shows that, in the collective case far from the boundary, N_s can be large, whereas in real polymers, intrachain rigidity can exceed the rigidity of substrate ten times but not ten thousand times. It follows that even the necessary condition of correctness of the Frenkel–Kontorova model requires nontrivially checking it for every real polymeric crystal.

In addition, in a polymeric crystal with all free chains, collective (involving many chains) phonon modes exist which have very narrow if any frequency gaps, and the lower boundary of phase velocity equals or approaches zero. Further, because any polymeric chain consists of several atomic rows linked with each other, the $\omega = \omega(k)$ dispersion curves (ω and k are the phonon frequency and wave vector) even of an isolated chain can be bent downward in the $k \longrightarrow 2\pi/c$ region (c is the chain period). Such dispersion curves also have no lower bound of phase velocities. Note that the existence of these modes is a direct consequence of the three-dimensional character of real polymeric crystals. A kink always emits radiation into these modes by the Vavilov–Cherenkov-type mechanism (when the velocity of the kink coincides with the phase velocity of some phonon mode), but the intensity of this radiation depends on the strength of coupling between the kink and such modes. For instance, numerical simulation of the dynamics of vacancies in zigzag chains of polyethylene crystals [16] for model [25] with united atoms in place of CH_2 groups shows that this radiation is only noticeable at velocities $v_k \gtrsim 0.6v_s$, and even at such velocities, its intensity is low.

5. AN EXAMPLE OF A CRYSTAL WHOSE DYNAMICS OF VACANCIES CAN BE DESCRIBED BY THE FRENKEL–KONTOROVA ONE-DIMENSIONAL MODEL

We have shown that, there exists a simple criterion for determining the amplitude and form of the potential generated by a row of atoms. This criterion is the $\delta = c/2r_0$ parameter value relating intrachain distances to the optimal r_0 distances of interchain interactions. At $\delta < 0.4$–0.55 (the collective case), this periodic potential is close to purely sinusoidal.

An arbitrary polymeric chain consists of several rows of almost rigidly fixed atoms, and the potential of such a chain can be obtained by summing the contributions of these rows. If the $\delta < 0.4$–0.55 condition is satisfied, each such contribution has a simple analytic form, which is important for calculating crystal unit cell parameters and the form and parameters of the interchain interaction potential.

Remarkably, in the simplest model of zigzag polyethylene chains [16, 25] with united atoms in place of CH_2 groups, the δ parameter for each row constituting a zigzag approximately equals 0.299. In a more realistic complete polyethylene model [26], which well reproduces the density, structure, and unit cell parameters of the orthorhombic polyethylene phase at room temperature, the δ parameters for carbon–carbon, hydrogen–hydrogen, and carbon–hydrogen interactions between rows are

$$\delta_{C-C} \approx 0.31, \quad \delta_{H-H} \approx 0.48, \quad \delta_{C-H} \approx 0.38,$$

which also falls into the category of collective interactions (although δ_{H-H} is close to the boundary value).

On the other hand, for a kink to emit no phonons into neighboring mobile crystal chains, it is necessary that the number of atoms over the $2L_k$ kink width be much larger than the N_s number of atoms that form substrate at the given chain point. In the collective case, $N_s \geq 4$. It follows that the $L_k \gg N_s/2 \geq 2$ inequality is the necessary condition for applying the continual (rather than discrete) sine-Gordon equation and for ignoring mobility of neighboring chains in describing the nonlinear dynamics of a chain with a kink.

For instance, in the same polyethylene model [16, 25, 26], $L_k \approx 16$ and $N_s \approx 4$ (for $\delta \sim 0.3$–0.5). As polyethylene parameters are at the boundary of the collective case, the rigidity of its chains is sufficiently large for the necessary condition $L_k \approx 16 \gg N_s/2 \approx 2$ to be satisfied.

We, however, know that this is not the sufficient condition. Molecular dynamics studies of this model [16] show that, at high ($v_k \gtrsim 0.6v_s$) kink velocities, there appears weak Vavilov–Cherenkov-type radiation to collective phonon modes of the crystal, the existence of which is a direct consequence of the three-dimensional character of real physical systems.

We nevertheless see that there exists at least one polymeric crystal, polyethylene, the behavior of vacancies in which can, with caution (at not very high velocities $v_k < 0.6v_s$), be described by the Frenkel–Kontorova model (using the approximation of immobile neighboring chains plus the continual sine-Gordon equation for the nonlinear dynamics of the chain under consideration).

6. CONCLUSION

The type of the dynamic behavior of a vacancy in a polymeric crystal is determined by the N_s number of neighboring chain atoms actually forming the substrate potential near each atom of the chain containing the defect. If this number is small ($N_s < 4$), the vacancy experiences deceleration and transmits energy to atoms of the chain containing it. If N_s is larger than the $2L_k$ vacancy width, the vacancy also experiences deceleration and transmits energy to atoms of the nearest neighboring chains. Only if the condition

$$4 \le N_s \ll 2L_k$$

is satisfied, the type of the dynamic behavior of vacancies is close to that of solitons, although, because of the three-dimensional character of systems, the energy of the defect is always slowly (in comparison with the characteristic period of chain vibrations) emitted into collective crystal phonon modes by the Vavilov–Cherenkov-type mechanism.

To summarize, we studied the possibility of applying the one-dimensional nonlinear integrable model to describe the dynamics of a topological localized soliton-type excitation in a real physical three-dimensional system. We found that there existed an interval of three-dimensional system parameters in which the one-dimensional model correctly predicted the soliton type of the dynamic behavior of defects. On the other hand, at other parameter values, defects, as distinguished from solitons, could not retain a constant velocity.

ACKNOWLEDGMENTS

The author thanks Prof. L.I. Manevich for support and fruitful discussions and the referee for valuable comments. The Mapple package was used to obtain numerical estimates and to construct plots. This work was financially supported by the Russian Foundation for Basic Research (projects nos. 01-03-33122 and 00-15-97431) and the Commission for Work with Young Scientists of the Russian Academy of Sciences (grant 123 of the 6th expert competition, 1999).

REFERENCES

1. M. S. Mansfield, Chem. Phys. Lett. **69**, 383 (1980).
2. J. L. Skinner and P. G. Wolynes, J. Chem. Phys. **73**, 4015 (1980); J. L. Skinner and P. G. Wolynes, J. Chem. Phys. **73**, 4022 (1980).
3. V. V. Ginzburg, L. I. Manevich, and N. G. Ryvkina, Mekh. Kompoz. Mater., No. 2, 249 (1991).
4. R. H. Boyd, Polymer **26**, 323 (1985); R. H. Boyd, Polymer **26**, 1123 (1985).
5. J. L. Skinner and Y. H. Park, Macromolecules **17**, 1735 (1984).
6. K. J. Wahlstrand, J. Chem. Phys. **82**, 5247 (1985); K. J. Wahlstrand and P. G. Wolynes, J. Chem. Phys. **82**, 5259 (1985); K. J. Wahlstrand, Polymer **29**, 256 (1988); K. J. Wahlstrand, Polymer **29**, 263 (1988).
7. J.-L. Syi and M. L. Mansfield, Polymer **29**, 987 (1988).
8. F. Zhang and M. A. Collins, Phys. Rev. E **49**, 5804 (1994); F. Zhang, M. A. Collins, and Y. S. Kivshar, Phys. Rev. E **51**, 3774 (1995).
9. A. V. Savin and L. I. Manevich, Phys. Rev. B **58**, 11386 (1998).
10. D. Bazeia and E. Ventura, Chem. Phys. Lett. **303**, 341 (1999).
11. R. L. McCullough and J. J. Hermans, J. Chem. Phys. **45**, 1941 (1966).
12. R. L. McCullough and J. M. Peterson, J. Appl. Phys. **44**, 1224 (1973).
13. T. Yemni and R. L. McCullough, J. Polym. Sci., Polym. Phys. Ed. **11**, 1385 (1973).
14. A. J. Hopfinger and S. K. Tripathy, CRC Crit. Rev. Solid State Mater. Sci. **9**, 285 (1980).
15. F. Zhang and M. A. Collins, Chem. Phys. Lett. **214**, 459 (1993).
16. E. A. Zubova, N. K. Balabaev, and L. I. Manevich, Zh. Éksp. Teor. Fiz. **115**, 1063 (1999) [JETP **88**, 586 (1999)].
17. E. A. Zubova, N. K. Balabaev, L. I. Manevich, and A. A. Tsygurov, Zh. Éksp. Teor. Fiz. **118**, 592 (2000) [JETP **91**, 515 (2000)].
18. O. M. Braun, Yu. S. Kivshar, and A. M. Kosevich, J. Phys. C **21**, 3881 (1988).
19. M. Peyrard and M. Remoissenet, Phys. Rev. B **26**, 2886 (1982).
20. M. Peyrard and M. D. Kruskal, Physica D (Amsterdam) **14**, 88 (1984).
21. M. B. Mineev, G. S. Mkrtchyan, and V. V. Schmidt, J. Low Temp. Phys. **45**, 497 (1981).
22. E. Goldobin, A. Wallraff, N. Tsyssen, and A. V. Ustinov, Phys. Rev. B **57**, 130 (1998).
23. R. Kleiner, T. Gaber, and G. Hechtfisher, Phys. Rev. B **62**, 4086 (2000).
24. J. F. Currie, S. E. Trullinger, A. R. Bishop, and J. A. Krumhansl, Phys. Rev. B **15**, 5567 (1977).
25. N. K. Balabaev, O. V. Gendel'man, M. A. Mazo, and L. I. Manevich, Zh. Fiz. Khim. **69**, 24 (1995).
26. E. Oleinik, I. Karmilov, S. Shenogin, *et al.*, Macromol. Symp. **146**, 133 (1999); Polimer Sci. A **42**, 1212 (2000).

Translated by V. Sipachev