Journal of Experimental and Theoretical Physics, Vol. 99, No. 3, 2004, pp. 449–459.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 3, 2004, pp. 515–526.

GRAVITATION, ASTROPHYSICS

Quasistatic Waves of Hydrogravity Generated in the Galactic Interstellar Medium by a Pulsating Neutron Star¶

S. Bastrukov[a,c,*], I. Molodtsova[a], D. Podgainy[a], V. Papoyan[a,b], J. Yang[c,d], and D. Murray[e]

[a] *Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia*
**e-mail: bast@jinr.ru*
[b] *Yerevan State University, Yerevan, 375025 Armenia*
[c] *Ewha Womans University 120-750, Seoul, South Korea*
[d] *Center for High Energy Physics 702-701, Daegu, South Korea*
[e] *Okanagan University College 3333, Kelowna, British Columbia, Canada*

Received February 25, 2004

Abstract—Based on principles of classical hydrodynamics and Newtonian gravity, the theory of hydrogravity, formulated in the manner of hydromagnetic theory, is developed to account for the gravitational effect of global pulsations of a star on the motions of the ambient gas–dust interstellar medium. Analytic derivation of the dispersion relation for canonical gravity waves at the free surface of an incompressible in viscid liquid is presented, illustrating practical usefulness of the proposed approach, heavily relying on the concept of classical gravitational stress introduced long ago by Fock and Chandrasekhar, and accentuating the shear character of this mode. Particular attention is given to gas-dynamical oscillations of a similar physical nature generated by a pulsating neutron star in an unbounded spherical shell of gas and dust promoted by circumstellar gravitational stresses and damped by viscosity of the interstellar matter. Computed in the long-wavelength approximation, the periods of these gravity-driven shear modes, referred to as quasistatic modes of hydrogravity, are found to be proportional to periods of the gravity modes in the neutron star bulk. Given that collective oscillations of cosmic plasma in the wave under consideration should be accompanied by electromagnetic radiation and taking into account that only the radio waves of this radiation can freely travel through the galactic gas–dust clouds, it is conjectured that the considered effect of gravitational coupling between seismic vibrations of a neutron star and fluctuations of the galactic interstellar medium should manifest itself in the radio range of pulsar spectra. Some useful implications of the theory developed here to a number of current problems of asteroseismology are briefly discussed.

1. INTRODUCTION

It was realized long ago that the restless behavior of neutron stars, exhibited in pulsar spectra by millisecond micropulses, owes its origin to seismic vibrations triggered either by implosion effects of supernova events or by starquakes [1–3], which may be connected with some short gamma-ray bursts [4, 5]. At present, there are tolerably coherent arguments showing that neutron stars (both pulsars and magnetars) can support long-lasting pulsations driven by bulk forces of elasticity, gravity, and magnetism of neutron star matter [6–10]. At the same time, the influence of neutron star pulsations on a gas–dust interstellar medium (ISM), which serves as a fluid matrix mediating a vast variety of gas-dynamical processes (e.g., [11]), remains less studied. This work discusses the hydrodynamic mechanism of gravitational coupling between seismic vibrations of a neutron star and fluctuations of gas–dust flows in the ambient envelope. Specifically, we consider a model in which a pulsating neutron star embedded in a gas–dust spherical shell is regarded as a source of large-scale hydrodynamical wave motions promoted by circumstellar gravitational stresses and damped by viscous stresses. The oscillatory motions in question have the same physical nature as the gravity waves at the free surface of an incompressible viscous fluid caused by the presence of a constant field of Newtonian gravity (e.g., [12–14]), the wave process being well known in the physics of planetary atmospheres [15].

This paper presents arguments that proper mathematical treatment of these gravity-driven wave motions of the interstellar medium can be developed on the basis of self-consistent equations for variables of classical hydrodynamics and Newtonian gravity, which are formulated in a manner of equations that govern the hydromagnetic theory. In pursuit of this aim, we follow two different approaches, both relying on the key concept of Newtonian gravitational stress. The underlying

¶This article was submitted by the authors in English.

idea of the first method, constituting the content of Sections 2 and 3, is to include the Newtonian gravitational field in a set of gas-dynamical variables of circumstellar motions by considering this field on an equal footing with the standard hydrodynamical variables such as density and velocity. The second method, formulated in the Appendix, is based on coupled equations involving density, the velocity, and the gravitational stress tensor. Particular attention is drawn to the fact that both these methods yield analytically identical estimates for the frequency and lifetime of the gravity modes owing their existence to fluctuations in circumstellar gravitational stresses caused by pulsations of a neutron star. In the discussion, we point out some useful applications of the theory developed here.

2. GOVERNING EQUATIONS OF HYDROGRAVITY

The point of departure in our considerations is the representation of the body force of gravity

$$\mathbf{F} = -\rho\mathbf{g}, \quad \nabla\mathbf{g} = 4\pi G\rho, \quad \mathbf{g} = -\nabla U \tag{1}$$

via the tensor of gravitational stresses G_{ik}:

$$F_i = -\rho g_i = -\frac{\partial G_{ik}}{\partial x_k},$$
$$G_{ik} = \frac{1}{4\pi G}\left[g_i g_k - \frac{1}{2}(g_j g_j)\delta_{ik}\right]. \tag{2}$$

To the best of our knowledge, this form of the gravitational force in the stationary material continuum of density ρ was first discussed long ago by Fock [17] and justified by Chandrasekhar [18]. Such a possibility is apparent from the identity

$$F_i = \rho\frac{\partial U}{\partial x_i} = -\frac{\partial}{\partial x_k}$$
$$\times\left\{\frac{1}{4\pi G}\left[\frac{\partial U}{\partial x_i}\frac{\partial U}{\partial x_k} - \frac{1}{2}\left(\frac{\partial U}{\partial x_j}\frac{\partial U}{\partial x_j}\right)\delta_{ik}\right]\right\}. \tag{3}$$

Also, the discussion of Newtonian gravitational stresses can be found in [19, 20]. A matter of particular interest for our present discussion is Chandrasekhar's suggestion [18] that the above tensor representation for the static force of Newtonian gravity be incorporated in the dynamical description of the gravity-driven motions of an inviscid fluid. Specifically, it is shown in [18] that replacement of the standard expression for the gravitational force,

$$F_i = -\rho g_i,$$

in the Euler equation for an ideal fluid

$$\rho\frac{dV_i}{dt} = -\frac{\partial P}{\partial x_i} - \rho g_i, \quad \frac{d}{dt} = \frac{\partial}{\partial t} + V_k\frac{\partial}{\partial x_k}, \tag{4}$$

by the above tensor representation

$$F_i = -\nabla_k G_{ik}$$

allows one to rewrite (4) in the form of a conservation law for the density of linear momentum ρV_i,

$$\frac{\partial(\rho V_i)}{\partial t} = -\frac{\partial\mathcal{P}_{ik}}{\partial x_k}, \quad \mathcal{P}_{ik} = \rho V_i V_k + P\delta_{ik} + G_{ik}, \tag{5}$$

where $\mathcal{P}_{ik}$ is the total flux density.

We recall at this point that the key statement in the MHD theory is that the state of motion of a magnetoactive fluid can be uniquely specified by the density $\rho(\mathbf{r}, t)$, the flow velocity $\mathbf{V}(\mathbf{r}, t)$, and the magnetic flux density $\mathbf{B}(\mathbf{r}, t)$, which are regarded on an equal footing as independent dynamical variables (see, e.g., [21–23]). The equations of dissipation-free MHD theory

$$\frac{\partial\rho}{\partial t} = -\frac{\partial\rho V_k}{\partial x_k},$$

$$\rho\frac{dV_i}{dt} = -\frac{\partial}{\partial x_i}\left(P + \frac{B^2}{8\pi}\right) + \frac{\partial}{\partial x_k}\left\{\frac{1}{8\pi}[B_i B_k + B_k B_i]\right\},$$

$$\frac{\partial B_i}{\partial t} = \frac{\partial}{\partial x_k}[V_i B_k - V_k B_i]$$

describe the fluid mechanics of a highly ionized (perfectly conducting) ISM threaded by a galactic magnetic field B.

Remarkably, Eq. (5) permits the equivalent representation

$$\rho\frac{dV_i}{dt} = -\frac{\partial}{\partial x_i}\left(P - \frac{g^2}{8\pi G}\right)$$
$$-\frac{\partial}{\partial x_k}\left\{\frac{1}{8\pi G}[g_i g_k + g_k g_i]\right\}, \tag{6}$$

which in appearance is similar to the Euler equation of the hydromagnetic model for interstellar gas dynamics. This then indicates that the constructive treatment of the gravity-driven gas dynamics of the interstellar medium can be developed on a methodological footing similar to that lying at the base of magnetohydrodynamics. In particular, this suggests that the gravitational field $\mathbf{g}(\mathbf{r}, t)$ can be regarded as an independent variable of the ISM motion on an equal footing with basic variables of interstellar gas dynamics, the density $\rho(\mathbf{r}, t)$ and the flow velocity $\mathbf{V}(\mathbf{r}, t)$. Then, adhering to this point of view, our next goal is to specify the form of the constitutive equation for the gravity-flow coupling, that is, an equation describing the kinematic relation between the vector field of classical gravity $\mathbf{g}(\mathbf{r}, t)$ and the density of linear momentum $\rho(\mathbf{r}, t)\mathbf{V}(\mathbf{r}, t)$.

It is customarily taken for granted that the time evolution of the density governed by the continuity equation

$$\frac{\partial \rho}{\partial t} = -\frac{\partial(\rho V_k)}{\partial x_k} \tag{7}$$

does not affect the analytic form of the equation for the static gravitational field,

$$\frac{\partial g_k}{\partial x_k} = 4\pi G\rho \longrightarrow \rho = \frac{1}{4\pi G}\frac{\partial g_k}{\partial x_k}. \tag{8}$$

The partial time derivative of the left-hand side of the last equation in (8) should therefore be equal to the left-hand side of continuity equation (7),

$$\begin{aligned}\frac{\partial \rho}{\partial t} &= \frac{\partial}{\partial x_k}\left(\frac{1}{4\pi G}\frac{\partial g_k}{\partial t}\right) = -\frac{\partial}{\partial x_k}(\rho V_k)\\ &\longrightarrow \frac{\partial}{\partial x_k}\left[\frac{\partial g_k}{\partial t} + 4\pi G\rho V_k\right] = 0.\end{aligned} \tag{9}$$

It follows that the equation

$$\frac{\partial g_k}{\partial t} = -4\pi G\rho V_k \tag{10}$$

resulting from the last identity is in agreement with both the equation of static gravity and the continuity equation. As postulated by the above arguments, Eq. (10) should be regarded as the constitutive equation for kinematic gravity-flow coupling. This shows that the standard equation for static Newtonian gravity,

$$\nabla_k g_k = 4\pi G\rho,$$

preserves its validity at all times, if valid initially. With all the above reservations in mind, we arrive at self-consistent equations of hydrodynamics and gravity-flow coupling,

$$\frac{d\rho}{dt} + \rho\frac{\partial V_k}{\partial x_k} = 0, \tag{11}$$

$$\begin{aligned}&\rho\frac{dV_i}{dt} + \frac{\partial}{\partial x_i}\left(P - \frac{g^2}{8\pi G}\right)\\ &+ \frac{\partial}{\partial x_k}\left\{\frac{1}{8\pi G}[g_i g_k + g_k g_i]\right\} = 0,\end{aligned} \tag{12}$$

$$\frac{\partial g_k}{\partial t} + 4\pi G\rho V_k = 0, \tag{13}$$

which in what follows are for brevity called the equations of hydrogravity.

The extension of these equations to the case of a viscous fluid is straightforward,

$$\begin{aligned}&\rho\frac{dV_i}{dt} + \frac{\partial W}{\partial x_i} - \frac{\partial P_{ik}}{\partial x_k} = \frac{\partial \Pi_{ik}}{\partial x_k}, \quad \Pi_{ik} = 2\nu\rho V_{ik},\\ &V_{ik} = \frac{1}{2}\left(\frac{\partial V_i}{\partial x_k} + \frac{\partial V_k}{\partial x_i}\right),\end{aligned} \tag{14}$$

$$W = P - \frac{g^2}{8\pi G}, \quad P_{ik} = -\frac{1}{8\pi G}[g_i g_k + g_k g_i], \tag{15}$$

where W is the total pressure and P_{ik} is the anisotropic gravitational stress tensor. We use Π_{ik} to denote the viscous stress tensor; V_{ik}, the rate-of-strain tensor; and ν, the kinematic viscosity of the gas–dust circumstellar medium.

3. HYDROGRAVITY MODES IN THE STELLAR COCOON MODEL

In the stellar cocoon model under consideration, a neutron star, embedded in a thick dusty shell of supernova debris, is regarded as a solid globe immersed in a spherical fluid matrix. To make the problem analytically tractable, we adopt the uniform density approximation for both the stellar matter ρ_s and the gas–dust medium ρ. It is presumed that the star and spherical dusty envelope are in hydrostatic equilibrium. This means that the characteristic times of the accretion processes are compared for a long time to the times of hydrodynamical fluctuations of gas–dust flow in the circumstellar shell. The purpose of this section is to delineate the equilibrium parameters of such an object, to wit, the spatial distribution of the static gravitational field $\mathbf{g}(r)$ and the hydrostatic pressure $P(r)$ in the regions of space relevant to the problem in question. In the absence of accretion processes, the above characteristics are determined by the equations

$$\begin{aligned}&\nabla\mathbf{g}_s(r) = 4\pi G\rho_s, \quad \nabla P_s(r) = -\rho_s\mathbf{g}_s(r),\\ &R < R_s,\end{aligned} \tag{16}$$

$$\begin{aligned}&\nabla\mathbf{g}^0(r) = 4\pi G\rho, \quad \nabla P_0(r) = -\rho\mathbf{g}^0(r),\\ &R_s < r < R.\end{aligned} \tag{17}$$

Equations (16) describe the static gravitational field $\mathbf{g}_s$ and hydrostatic pressure P_s in the star bulk, R_s is the star radius, and

$$M_s = (4\pi/3)\rho_s R_s^3$$

is the star mass. Equations (17) determine the static gravitational field $\mathbf{g}^0$ and hydrostatic pressure P_0 in the

dusty cocoon of radius R. In the star's interior, the solutions of (16) are given by

$$\mathbf{g}_s(r) = \frac{4\pi}{3}G\rho_s\mathbf{r}, \quad P_s(r) = \frac{2\pi}{3}\rho_s^2 G(R_s^2 - r^2), \quad r < R_s. \tag{18}$$

In the stellar envelope, the static gravitational field $\mathbf{g}^0(r)$ and hydrostatic pressure $P_0(r)$ are given by the following solutions of Eqs. (17):

$$\mathbf{g}^0(r) = \frac{4\pi}{3}G\rho\left[1 + \frac{(\rho_s - \rho)}{\rho}\frac{R_s^3}{r^3}\right]\mathbf{r}, \quad R_s < r < R, \tag{19}$$

$$P_0(r) = \frac{2\pi}{3}G\rho^2(R^2 - r^2) + \frac{4\pi}{3}G\frac{R_s^3}{R}\rho(\rho_s - \rho)\frac{R - r}{r}, \quad R_s < r < R. \tag{20}$$

The detailed derivation of this latter equation, in a somewhat different context, can be found in [16]. From this point on, superscript zero labels the static gravity field in the ambient gas–dust envelope.

3.1. Shear Oscillations Driven by Gravitational Stresses

In what follows, we consider small-amplitude disturbances in the dense gas–dust shell of a stellar envelope, generated by seismic vibrations of the central neutron star. We assume that these are not accompanied by fluctuations in density. While the model of incompressible viscous fluid is admittedly a highly idealized approximation, nevertheless we do not expect to lose any essential gas-dynamical features of gravity waves in the interstellar medium on this account. Under such disturbances, the quantities entering the equations of hydrogravity are infinitesimally perturbed as

$$V_i = V_i^0 + v_i, \quad V_i^0 = 0, \quad g_i = g_i^0 + \delta g_i, \quad P = P_0 + p, \tag{21}$$

$$W = W_0 + w, \quad W_0 = P_0 - \frac{g_0^2}{8\pi G}, \quad w = p - \frac{1}{4\pi}g_k^0\delta g_k, \tag{22}$$

$$P_{ik} = P_{ik}^0 + t_{ik}, \quad P_{ik}^0 = -\frac{1}{8\pi G}[g_i^0 g_k^0 + g_k^0 g_i^0], \quad t_{ik} = -\frac{1}{4\pi G}[g_i^0\delta g_k + g_k^0\delta g_i], \tag{23}$$

$$\Pi_{ik} = \Pi_{ik}^0 + \pi_{ik}, \quad \Pi_{ik}^0 = 0, \quad \pi_{ik} = 2\nu\rho v_{ik}, \quad v_{ik} = \frac{1}{2}\left(\frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i}\right), \tag{24}$$

where g_i^0 and P_0 are determined by Eqs. (19) and (20), respectively. Inserting (21)–(24) into (11)–(13), we arrive at the linearized equations of hydrogravity for heavy incompressible viscous fluid,

$$\rho\frac{\partial v_i}{\partial t} = -\frac{\partial w}{\partial x_i} + \frac{\partial t_{ik}}{\partial x_k} + \frac{\partial \pi_{ik}}{\partial x_k}, \quad \frac{\partial v_k}{\partial x_k} = 0, \tag{25}$$

$$\frac{\partial \delta g_i}{\partial t} = -4\pi G\rho v_i. \tag{26}$$

Scalar multiplication of (25) with v_i and integration over the cocoon volume (ignoring the effects of surface stresses) yields

$$\frac{\partial}{\partial t}\int\frac{\rho v^2}{2}d^3V = -\int[t_{ik} + \pi_{ik}]v_{ik}d^3V. \tag{27}$$

This equation gives the rate at which the kinetic energy of the gas-dynamical motions changes. The most important point for our present discussion is that gravitational forces in the volume of the dusty shell do work, which is characterized by the inseparable link between anisotropic gravitational stresses and shear fluctuations of material flow, $t_{ik}v_{ik}$. This suggests that gravitational stresses impart to the ambient gas–dust matter a portion of mechanical rigidity generic to viscoelastic materials whose response to an external disturbance is accompanied by shear fluctuations.

In appearance, Eq. (27) is similar to the equation of energy balance for shear response of an isotropic viscoelastic material continuum,

$$\frac{\partial}{\partial t}\int\frac{\rho\dot{u}^2}{2}d^3V = -\int[\sigma_{ik} + \pi_{ik}]\dot{u}_{ik}d^3V, \quad \sigma_{ik} = 2\mu u_{ik}, \quad u_{ik} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_k} + \frac{\partial u_k}{\partial x_i}\right), \quad \frac{\partial u_k}{\partial x_k} = 0,$$

where u_i is the material displacement field (related to the velocity as $v_i = \dot{u}_i$, which implies $v_{ik} = \dot{u}_{ik}$ for the rate of strains), σ_{ik} is the Hookean shear stress, u_{ik} is the shear strain tensor, and μ is the shear modulus. This last equation is obtained by scalar multiplication with $\dot{u}_i$ of the basic equation of continuum mechanics of viscoelastic incompressible matter,

$$\rho\ddot{u}_i = \nabla_k\sigma_{ik} + \nabla_k\pi_{ik},$$

followed by integration over the volume.

It is expected, therefore, that a pulsating neutron star should generate shear oscillations of gas–dust matter in the circumstellar envelope. Clearly, the only way of exploring this statement is to evaluate the frequency and lifetime of the mode promoted by gravitational stresses. In doing this, we use the energy variational principle. The procedure is as follows. The first step is to use the separable representation for fluctuating variables

$$v_i(\mathbf{r}, t) = a_i(\mathbf{r})\dot{\alpha}(t), \quad v_{ik}(\mathbf{r}, t) = a_{ik}(\mathbf{r})\dot{\alpha}(t), \qquad a_{ik} = \frac{1}{2}[\nabla_i a_k + \nabla_k a_i]. \tag{28}$$

Substituting this form for v_i in (26) and eliminating the time derivative, we obtain

$$\delta g_i(\mathbf{r}, t) = -4\pi G\rho a_i(\mathbf{r})\alpha(t). \tag{29}$$

The analogous separable forms for tensors of gravitational t_{ik} and viscous π_{ik} stresses are

$$t_{ik}(\mathbf{r}, t) = \tau_{ik}(\mathbf{r})\alpha(t), \qquad \tau_{ik}(\mathbf{r}) = \rho[g_i^0(\mathbf{r})a_k(\mathbf{r}) + g_k^0(\mathbf{r})a_i(\mathbf{r})], \qquad \pi_{ik} = 2\rho\nu a_{ik}\dot{\alpha}. \tag{30}$$

Hereafter, $a_i(\mathbf{r})$ is the instantaneous displacement field and $\alpha(t)$ is the temporal amplitude of the oscillations.

Inserting (28) and (30) in (27), we arrive at the equation governing the time evolution of $\alpha(t)$ having the form of the standard equation for a damped harmonic oscillator,

$$M\ddot{\alpha}(t) + D\dot{\alpha}(t) + K\alpha(t) = 0, \tag{31}$$

where the parameters (inertia M, stiffness K, and viscous friction D) are given by

$$M = \int \rho a_i(\mathbf{r})a_i(\mathbf{r})d^3V, \qquad D = \frac{1}{2}\int \rho\nu[\nabla_i a_k + \nabla_k a_i][\nabla_i a_k + \nabla_k a_i]d^3V, \tag{32}$$

$$K = \frac{1}{2}\int \rho[g_i^0 a_k + g_k^0 a_i][\nabla_i a_k + \nabla_k a_i]d^3V. \tag{33}$$

The well-known solution of (31) is

$$\alpha(t) = \alpha_0 \exp(-t/\tau)\cos\omega t,$$

where

$$\omega^2 = \omega_0^2[1 - (\omega_0\tau)^{-2}], \quad \omega_0^2 = \frac{K}{M}, \quad \tau = \frac{2M}{D}. \tag{34}$$

Here, ω_0 stands for the frequency of nondissipative free oscillations and τ is the viscous damping time. Long-lasting oscillations exist if $\omega_0\tau \gg 1$. Thus, to evaluate the frequency and lifetime of the gravity modes in the spherical gas–dust nebula surrounding a pulsating neutron star, we must specify the form of the instantaneous displacement $\mathbf{a}(\mathbf{r})$, which is solenoidal in view of our adopted approximation of incompressible fluid. In doing this, we consider the quasistatic regime of wave motions of a spherical envelope generated by seismic vibrations of a neutron star. In the case of quasistatic waves, the velocity field $\mathbf{v}(\mathbf{r}, t)$ is determined by the solutions of the Laplace equation

$$\nabla^2\mathbf{v}(\mathbf{r}, t) = 0,$$

which is regarded as the long-wavelength limit of the Helmholtz equation of standing waves

$$\nabla^2\mathbf{v}(\mathbf{r}, t) + k^2\mathbf{v}(\mathbf{r}, t) = 0,$$

since in the limit of extremely long wavelengths ($\lambda \longrightarrow \infty$), the wave vector $k = 2\pi/\lambda \longrightarrow 0$. In view of the relation

$$\mathbf{v} = \mathbf{a}(\mathbf{r})\dot{\alpha}(t),$$

the instantaneous displacement field $\mathbf{a}(\mathbf{r})$ satisfies the equations

$$\nabla^2\mathbf{a}(\mathbf{r}) = 0, \quad \nabla\mathbf{a}(\mathbf{r}) = 0. \tag{35}$$

The poloidal solution of (35),

$$\mathbf{a}_s(\mathbf{r}) = (N/l)[\nabla \times [\nabla \times \mathbf{r}r^{-l-1}P_l(\mu)]] = -N\nabla r^{-(l+1)}P_l(\mu), \quad \mu = \cos\theta, \tag{36}$$

describes a spheroidal quasistatic wave (the long-wavelength limit of a standing wave) in the circumstellar envelope. Here, $P_l(\mu)$ is the Legendre polynomial of degree l; the spherical polar coordinate system with fixed polar axis is used. This irrotational ($\nabla \times \mathbf{a}_s = 0$) field of displacement is generated by a spheroidally pulsating neutron star.

A neutron star executing global torsional vibrations produces fluctuations of gas–dust flows of substantially differentially-rotational character. The field of material displacements in this kind of oscillatory motions of the circumstellar shell is given by the toroidal vector field

$$\mathbf{a}_t(\mathbf{r}) = [\phi(\mathbf{r}) \times \mathbf{r}], \quad \phi = N\nabla(r^{-(l+1)}P_l(\mu)). \tag{37}$$

This field is also the general solution of (35). It is noteworthy that the parameters M, K, and D depend on arbitrary constant N as N^2, and hence, ω_0 and τ are independent of the specific form of N.

3.2. Spheroidal Hydrogravity Mode

We assume that R, the radius of the circumstellar cloud, is much larger than the star's radius R_s:

$$R_s/R \ll 1.$$

Therefore, the limits of integration along the radial variable r can be taken from the surface of the star, $r = R_s$, to the outer surface of the gas–dust shell removed to infinity, $r = R \longrightarrow \infty$; a neutron star looks like an oscil-

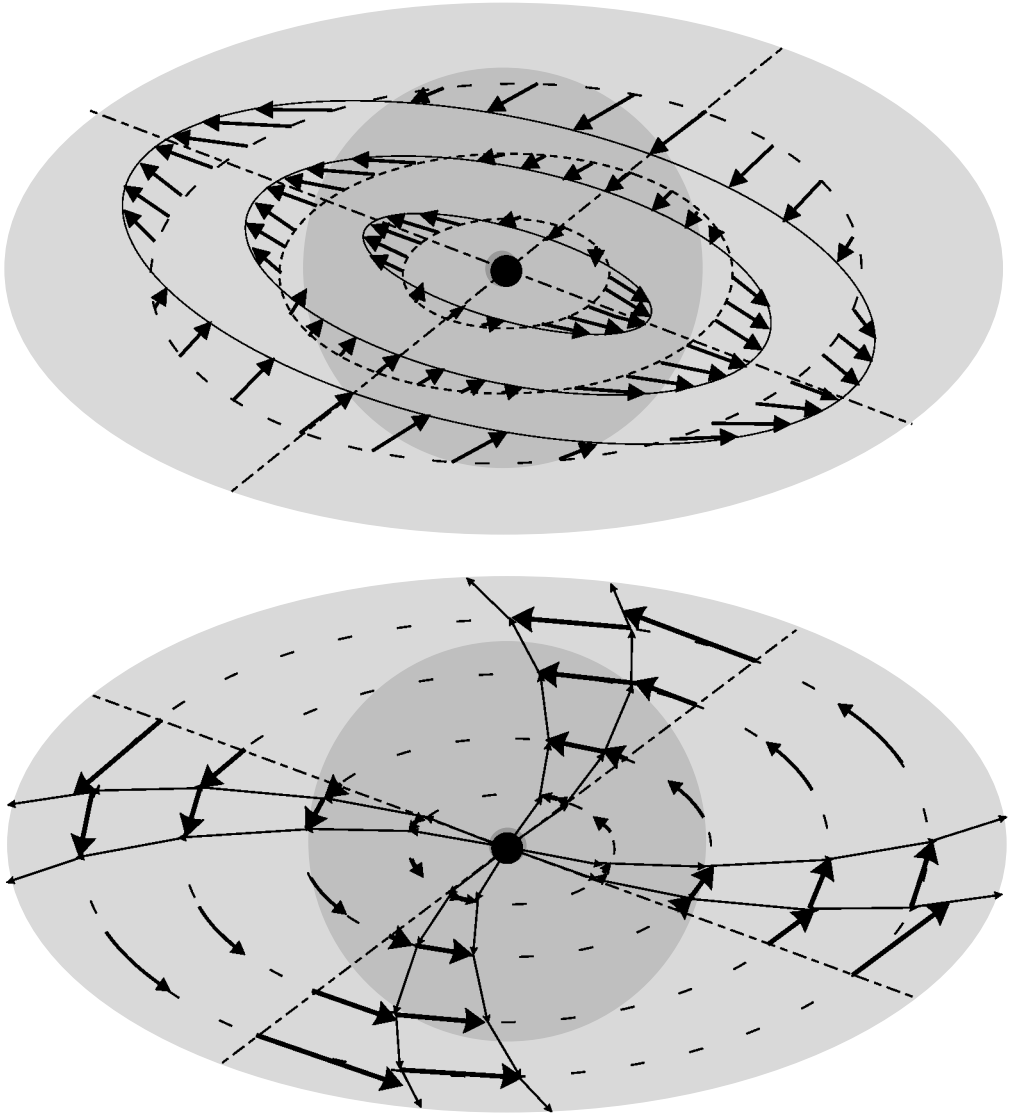

Fig. 1. Geometric illustration of material displacements in circumstellar hydrogravity waves generated by a spheroidally (top picture) and torsionally (bottom picture) oscillating neutron star.

lating blob immersed in a spherical gas–dust matrix of infinitely large radius. For the parameters of inertia $M_s(l)$, internal friction $D_s(l)$, and the lifetime $\tau_s(l)$ of a spheroidal *g*-mode, computed as a function of the multipole degree of oscillations (l), we obtain

$$M_s(l) = 4\pi\rho\frac{N^2}{R_s^{2l+1}}\frac{l+1}{2l+1},$$
$$D_s(l) = 8\pi\eta\frac{N^2}{R_s^{2l+3}}\frac{l+1}{l+2}, \quad (38)$$

$$\tau_s(l) = \tau[(2l+1)(l+2)]^{-1},$$
$$\tau = \frac{R_s^2}{\nu}, \quad \nu = \frac{\eta}{\rho}, \quad (39)$$

where η is the dynamical viscosity of the interstellar medium and τ is the time constant of exponential time decay due to viscous dissipation of the oscillatory motions. In somewhat different context, this last expression for $\tau_s(l)$ has first been established by Lamb [12]. More laborious are calculations of the restoring force parameter K of hydrogravity, whose analytic form is given by expression (33). We omit discussion of the tedious integration procedure and only note that these calculations are appreciably facilitated by the use of Maple symbolic algebra software. As a result, we obtain

$$K_s(l) = \frac{16\pi^2}{3}N^2\frac{G\rho^2}{R_s^{2l+1}}(l+1)\left[\frac{2(l+2)}{2l+1}+\frac{\rho_s-\rho}{\rho}\right]. \quad (40)$$

Figure 1 illustrates circumstellar gravitational stresses generated by spheroidal quadrupole vibrations of a neutron star. The frequency of undamped spheroidal modes of hydrogravity in the galactic ISM is given by

$$\omega_{0s}^2(l) = \frac{4\pi}{3}G\rho(2l+1)\left[\frac{2(l+2)}{2l+1}+\frac{\rho_s-\rho}{\rho}\right]. \quad (41)$$

Because the density of the star ρ_s is much greater than that of the ambient interstellar medium ρ, that is, $\rho/\rho_s \ll 1$, the last formula (41) can be replaced by

$$\omega_{0s}^2(l) = \omega_G^2(2l+1), \quad \omega_G^2 = \frac{4\pi}{3}G\rho_s = \frac{GM_s}{R_s^3}, \quad (42)$$

where M_s and R_s are the mass and radius of the neutron star and ω_G is the natural unit of frequency of *g*-modes in the star bulk. This result is perhaps the most striking outcome of the considered models, which shows that the frequency of the hydrogravity mode in the ISM is independent of the density of the galactic interstellar matter. Formula (42) can be compared with that for the frequency of nonradial spheroidal *g*-modes in the neutron star bulk computed in [24] as a function of the multipole degree *l*,

$$\omega_s^2({}_0G_l) = \omega_G^2[2(l-1)]. \quad (43)$$

We see that in the limit of very high overtones, $l \gg 1$, the frequency of spheroidal *g*-modes in the ISM coincides with that for *g*-modes in the neutron star bulk,

$$\omega_{0s}(l) = \omega_s({}_0G_l).$$

3.3. Toroidal Hydrogravity Mode

The inertia $M_t(l)$, internal friction $D_t(l)$, and lifetime $\tau_t(l)$ as functions of the multipole degree *l* of the toroidal hydrogravity mode are given by

$$M_t(l) = N^2\frac{4\pi\rho}{R_s^{2l-1}}\frac{l(l+1)}{(2l+1)(2l-1)},$$
$$D_t(l) = N^2\frac{4\pi\eta}{R_s^{2l+1}}\frac{l(l+1)(l+2)}{(2l+1)}, \quad (44)$$

$$\tau_t(l) = 2\tau[(2l-1)(l+2)]^{-1},$$
$$\tau = \frac{R_s^2}{\nu}, \quad \nu = \frac{\eta}{\rho}. \quad (45)$$

The obtained analytic estimate (45) for the time of viscous relaxation of rotational oscillations in the circumstellar envelope (pictured in the lower part of Fig. 1) has the same practical usefulness as Lamb's formula (39). It follows that high-multipole gravity modes decay faster then low-multipole ones, and this conclusion is independent of the adopted approximation of incompressible matter. An analogous conclusion was derived in [26] for the decay time of the toroidal mode in the neutron star bulk. In a cosmic hydrogen plasma,

$$\rho\nu = 2.2\times10^{-15}T^{5/2}/\ln\Lambda.$$

The numerical value of the factor $\Lambda \approx 10$–15 and $\rho \approx 10^{-24}$ g cm^{-3} [15]. For $\rho\nu \approx 10^{-15}$ cm^2 s, the time of viscous relaxation $\tau_\nu \approx 10^3$ s; for $\rho\nu = 10^{-20}$ cm^2 s, the lifetime of interstellar mode of hydrogravity is on the order of 10 years. Another way of computing the time of viscous dissipation is based on the formula

$$\eta = nk_BT/\nu$$

(where ν is the collision frequency in the gas–dust interstellar matter, *n* is the particle density, and *T* is the temperature), which leads to lifetimes from 10^2 to 10^6 years. All this leaves no doubt about the existence of the hydrogravity mode in the galactic interstellar medium.

The restoring force parameter $K_t(l)$ is given by

$$K_t(l) = \frac{8\pi^2}{3}N^2\frac{G\rho^2}{R_s^{2l-1}}\frac{l(l+2)}{2l+1}\left[\frac{2(l+1)}{2l-1}+\frac{\rho_s-\rho}{\rho}\right]. \quad (46)$$

The frequency of dissipationless toroidal hydrogravity mode in the ambient ISM is given by

$$\omega_{0t}^2(l) = \frac{2\pi}{3}G\rho\frac{(2l-1)(l+2)}{l+1}\left[\frac{2(l+1)}{2l-1}+\frac{\rho_s-\rho}{\rho}\right]. \quad (47)$$

In the natural limit $\rho/\rho_s \ll 1$, the latter formula is replaced by

$$\omega_{0t}^2(l) = \omega_G^2\frac{(2l-1)(l+2)}{2(l+1)}, \quad (48)$$

where ω_G stands for the frequency of *g*-modes in the neutron star bulk. This equation again highlights the fact that the frequency of the hydrogravity mode is independent of the material properties of galactic interstellar matter. For the sake of comparison, the frequency of the torsional *g* modes in the neutron star is given by [25]

$$\omega_t^2({}_0G_l) = \omega_G^2(l-1). \quad (49)$$

We see that in the limit $l \gg 1$, we have $\omega_{0t}(l) = \omega_t({}_0G_l)$. The difference between periods of hydrogravity modes in the ISM and *g* modes in the neutron star bulk is illustrated in Fig. 2. Our expectation that the kind of inter-

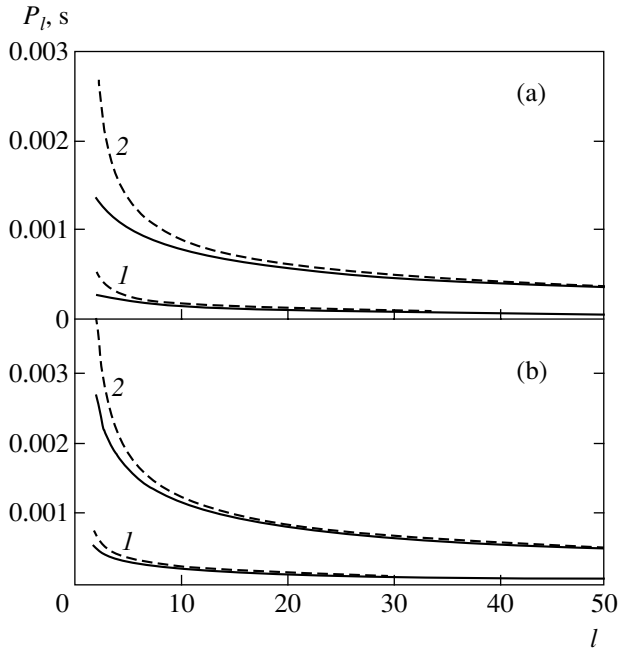

Fig. 2. Period P_l, in seconds, as a function of multipole degree l of spheroidal (a) and torsional (b) hydrogravity modes in the circumstellar envelope (solid line) and the corresponding g modes in the bulk of a neutron star (dashed line) computed for two models of neutron stars. *1*—$M_S = 1.3M_\odot$, $R_S = 13$ km; *2*—$M_S = 0.1M_\odot$, $R_S = 18$ km.

stellar motions considered can be detected in an electromagnetic signal from a pulsating neutron star rests on the plausible assumption that collective oscillations of charged species in interstellar gravity waves should be accompanied by emission of electromagnetic waves. Since the radio range of such an emission is not extinguished by the gas–dust cloud of the ISM, the interstellar gravity waves could manifest themselves by a periodic radio signal whose timing is determined by the frequency of the gravity mode in the central star. Also, it is noteworthy that in [27], it is shown that taking into account electromagnetic processes around a torsionally oscillating neutron star can lead to a better understanding of the behavior of gamma-ray bursts.

3.4. Canonical Gravity Waves from Equations of Hydrogravity

In this subsection, we show that the approach suggested above regains the well-known results of the classical fluid-dynamic theory regarding the dispersion relation for the free surface gravity waves in an incompressible inviscid liquid caused by the presence of a constant gravitational field (see, e.g., [12–14]). This is interesting in its own right because the developed treatment discloses the fact that the classical gravity waves are of a substantially shear character. From the energy balance equation (27) obtained above, it follows that the dissipative free fluctuations of an incompressible liquid promoted by anisotropic Newtonian gravitational stresses are governed by equations of the form

$$\rho\dot{v}_i = -(4\pi G)^{-1}\nabla_k[g_i^0\delta g_k + g_k^0\delta g_i], \qquad \delta\dot{g}_i = -4\pi G\rho v_i. \tag{50}$$

These equations should be supplemented by the incompressibility condition

$$\nabla\mathbf{v} = 0,$$

which implies that

$$\nabla\delta\mathbf{g} = 0.$$

Adhering to the treatment of hydrodynamic gravity waves in an incompressible fluid of infinite depth, given in [14], we assume that the plane $z = 0$ corresponds to the equilibrium fluid surface. In this case, the constant gravitational field $\mathbf{g}^0$ has the components

$$\mathbf{g}^0 = [g_x = 0, g_y = 0, g_z = -g] = \text{const}. \tag{51}$$

The motions are restricted to the *xz* plane, which means that the fluctuating field of the velocity is a function of x and z and therefore has just two nonzero components

$$v_x = v_x(x, z), \quad v_y = 0, \quad v_z = v_z(x, z). \tag{52}$$

Given (51) and (52), in Cartesian coordinates, Eqs. (50) break up into the set of noncombining equations

$$\rho\frac{\partial v_x}{\partial t} = \frac{g}{4\pi G}\frac{\partial\delta g_x}{\partial z}, \quad \frac{\partial\delta g_x}{\partial t} = -4\pi G\rho v_x, \tag{53}$$

$$\rho\frac{\partial v_z}{\partial t} = \frac{g}{4\pi G}\frac{\partial\delta g_z}{\partial z}, \quad \frac{\partial\delta g_z}{\partial t} = -4\pi G\rho v_z, \tag{54}$$

exhibiting strong coupling between fluctuations in the velocity of hydrodynamical flow and gravitational field. Taking the time derivative in Eqs. (53) and (54), we find that the resultant equations can easily be combined to give identical equations for each fluctuating variable,

$$\frac{\partial^2 v_x}{\partial t^2} = -g\frac{\partial v_x}{\partial z}, \quad \frac{\partial^2\delta g_x}{\partial t^2} = -g\frac{\partial\delta g_x}{\partial z}, \tag{55}$$

$$\frac{\partial^2 v_z}{\partial t^2} = -g\frac{\partial v_z}{\partial z}, \quad \frac{\partial^2\delta g_z}{\partial t^2} = -g\frac{\partial\delta g_z}{\partial z}. \tag{56}$$

We consider an incompressible liquid with a free surface. A free gravity wave is interpreted as a disturbance traveling in this liquid whose amplitude is exponentially decreasing toward the depth, i.e., as $z \longrightarrow -\infty$, and for which the velocity components are described by

$$v_x = -Ake^{kz}\sin(kx - \omega t), \quad v_y = 0, \qquad v_z = Ake^{kz}\cos(kx - \omega t). \tag{57}$$

The components of the fluctuating gravity field in this

wave have the form

$$\delta g_x = \frac{4\pi\rho}{\omega} v_z, \quad \delta g_y = 0, \quad \delta g_z = -\frac{4\pi\rho}{\omega} v_x,$$

where v_x and v_z are given by expressions (57). Substituting (57) in (55), we arrive at the well-known dispersion relation of a classical free surface gravity wave in an incompressible liquid of infinite depth [14, 15],

$$\omega = \sqrt{gk}, \quad V_G = \frac{\partial\omega}{\partial k} = \frac{1}{2}\sqrt{\frac{g}{k}}, \tag{58}$$

where V_G is the group velocity of the gravity wave. In this wave, the velocity vector **v**, at any fixed value of the depth coordinate *z* in the *xz* plane, undergoes a uniform rotation in this plane preserving its magnitude. This is clearly seen by representing the flow velocity as

$$\begin{aligned} \mathbf{v} &= [\nabla\times(\mathbf{e}_y f)] = [\nabla f\times\mathbf{e}_y], \\ f &= Ae^{kz}\sin(kx-\omega t), \end{aligned} \tag{59}$$

where $\mathbf{e}_y$ is the unit vector in the positive direction of the *y* axis around which, at a fixed value of *z*, the velocity vector **v** executes uniform rotation. However, it should be clearly realized that this rotation has nothing to do with vorticity of the fluid flow, since the vector field of vorticity

$$\boldsymbol{\Omega} = \nabla\times\mathbf{v} = 0.$$

On the other hand, the requirements

$$\nabla\cdot\mathbf{v} = 0, \quad \nabla\times\mathbf{v} = 0 \tag{60}$$

imply that **v** can be represented as gradient of a scalar function ϕ

$$\mathbf{v} = \nabla\phi, \quad \phi = Ae^{kz}\cos(kx-\omega t). \tag{61}$$

It can be verified that the velocity field in the surface gravity wave **v**, Eq. (57), obeys the equation

$$\nabla^2\mathbf{v} = 0$$

as well. It is this fact that has been used as a guide in the above adopted classification of the gravity modes in a spherical circumstellar shell as spheroidal and toroidal modes in which the fields of displacements are described by two general solutions to the vector Laplace equation. Thus, the proposed equations of hydrogravity provide a proper account of the canonical gravity waves by accentuating the shear character of this mode. Essentially, this means that gravitational stress endows an incompressible fluid with the mechanical properties typical of viscoelastic materials capable of transmitting shear waves.

4. DISCUSSION

An understanding of physical mechanisms governing the large-scale motions of galactic interstellar medium brought about by seismic vibrations of stars is important in two areas of current astrophysics: asteroseismology and interstellar gas dynamics. In this work, we have investigated the wave motions of the galactic interstellar medium promoted by circumstellar gravitational fields of pulsating stars. In doing this, we have set up self-consistent equations of hydrogravity having in appearance some features in common with those lying at the base of hydromagnetic theory. By examining the potential of such an approach, heavily relying on the concept of Newtonian gravitational stresses, we have shown that the proposed theory regains the dispersion equation for the canonical gravity waves traveling near the surface of an incompressible inviscid liquid of infinite depth, the wave process being well known in the theoretical oceanology and physics of planetary atmospheres. Newly highlighted here is the shear character of oscillating flows in this wave, owing its origin to fluctuations of Newtonian gravitational stresses.

Based on this and working from the homogeneous model of a spherical stellar cocoon (a star surrounded by an extended spherical shell of a gas–dust medium), we apply the proposed theory of hydrogravity to analysis of the small-amplitude gravity modes generated in the interstellar medium by a neutron star executing spheroidal and torsional vibrations in a quiescent, presumably poststarquake, regime. In the calculations presented (carried out by two constructively different operational tools), the approximation of incompressible viscous fluid has been adopted. This implies that disturbances coming from a pulsating neutron star lead to weak perturbations accompanied by coupled fluctuations of the velocity and gravitational field, whereas the equilibrium density and hydrostatic pressure in the ambient gas–dust shell remain unaffected. Clearly, such an approximation is unwarranted for violent starquakes generating shock and compressional waves. The extension of the proposed theory to the case of these latter waves requires special investigation, which is outside the scope of our present discussion. The practical usefulness of the considered, admittedly idealized, model is that it allows one to attain conclusive inferences regarding the dependence of period and lifetime of considered modes upon characteristic parameters of both a pulsating star and surrounding gas–dust interstellar matter and the multipole degree of oscillations as well. From the physical side, a finding of particular interest is that the frequency of these weakly attenuated modes of hydrogravity in an unbounded dusty envelope is proportional to the frequency of the *g* mode in the bulk of a neutron star. The corresponding period falls in the interval from 0.1 to 20 milliseconds. This inference is in agreement with the Boriakoff–Van Horn conjecture [2, 3] that micropulses of millisecond duration clearly discernible in the windows of the main pulse train owe their existence to pulsations of neutron stars. Together with this, it seems fairly plausible that unpredictable irregularities and perturbations in the ISM mediating the waves of hydrogravity under consider-

ation should substantially affect the coherency of these micropulses, as has been extensively discussed in [28].

While the developed theory is presented in the context of neutron star pulsations, it is hoped that the theoretical predictions inferred here can find useful applications in the area of asteroseismology. In this connection, it is important that the considered mechanisms of gravitational coupling between small-amplitude vibrations of a star and a stellar envelope presume that collective oscillations of charged particles (forming circumstellar plasma) in the quasistatic wave of hydrogravity are accompanied by electromagnetic radiation whose frequency coincides with that for the gravity mode in the star bulk. It is expected, therefore, that such radiation can be observed in the vicinity of any star surrounded by interstellar plasma and executing small-amplitude nonradial vibrations driven by self-gravity. In this case, the considered quasistatic waves of hydrogravity can exist in the solar envelope, provided the Sun undergoes global nonradial gravity-driven vibrations of small amplitude with frequencies proportional to the basic frequency of g mode ω_G. Such an attitude sheds some new light on the known problem of helioseismology [29] regarding 160-min variability discovered long ago in solar observations [30, 31], which has been interpreted as a manifestation of the gravity-driven vibrations of the Sun (see also [32]). However, in subsequent years, the authenticity of the solar origin of this signal has been the subject of controversy (e.g., [33]). In recent work [34], advocating the solar origin of this signal, it is argued that this variability cannot be ascribed to some terrestrial cause or to an artifact of the data reduction procedure. Notwithstanding the fact that further measurements (preferably with the use of satellite-based telescopes) are needed to attain more definite statements regarding the very source and physical nature of this intriguing signal, we conclude that predictions of the theory developed in this work are in line with the hypothesis on the helioseismic origin of this phenomenon.

ACKNOWLEDGMENTS

We are indebted to M.G. Park, J.C. Hwang, D. Mkrtichian, C.H. Lee, G.S. Bisnovatyi-Kogan, M. van Putten, H. Noh, and H.K. Lee for discussion and valuable comments. The authors acknowledge partial support of this work by the Eastern Europe exchange program (grant no. 02 Russia 10-58-2) sponsored by KISTEP, South Korea, and the Russian Foundation for Basic Research (project no. 02-01-00606).

APPENDIX

We show that the results obtained above can be derived on a different mathematical footing. The basic idea of this method is to use the gravitational stress tensor G_{ik} as the dynamical variable of motions together with the density ρ and the velocity field V_i, the evolution of which is governed by coupled equations of the form

$$\frac{d\rho}{dt} + \rho\frac{\partial V_k}{\partial x_k} = 0, \tag{62}$$

$$\rho\frac{dV_i}{dt} + \frac{\partial P}{\partial x_i} - \frac{\partial G_{ik}}{\partial x_k} = \frac{\partial \Pi_{ik}}{\partial x_k}, \tag{63}$$

$$\frac{dG_{ik}}{dt} + G_{ij}\frac{\partial V_k}{\partial x_j} + G_{kj}\frac{\partial V_i}{\partial x_j} + G_{ik}\frac{\partial V_j}{\partial x_j} = 0. \tag{64}$$

Such an approach has been used in [8, 9] to compute the gravity modes in the bulk of a neutron star.

As in the previous section, we focus on disturbances of the gas–dust cocoon triggered by seismic vibrations of a neutron star that are not accompanied by fluctuations in density but solely in velocity, pressure, and gravitational stresses,

$$\begin{aligned} V_k(\mathbf{r}, t) &= V_k^0(0) + v_k(\mathbf{r}, t), \\ P(\mathbf{r}, t) &= P_0(r) + P(\mathbf{r}, t), \\ G_{ik}(\mathbf{r}, t) &= G_{ik}^0(r) + g_{ik}(\mathbf{r}, t). \end{aligned} \tag{65}$$

We note that in this model, the static gravitational stresses in the stationary cloud surrounding the neutron star are determined by the hydrostatic pressure (Eq. (20)) as

$$\begin{aligned} G_{ik}(r) &= -P_0(r)\delta_{ik}, \\ P_0(r) &= \frac{2\pi}{3}G\rho^2(R_s^2 - r^2) \\ &+ \frac{4\pi}{2}G\rho(\rho_s - \rho)R_s^3\left(\frac{1}{r} - \frac{1}{R_s}\right). \end{aligned} \tag{66}$$

Inserting (65) in (62)–(64), we arrive at linearized equations of gravity-driven fluctuations,

$$\rho\frac{\partial v_i}{\partial t} = -\frac{\partial \rho}{\partial x_k} + \frac{\partial g_{ik}}{\partial x_k} + \frac{\partial \pi_{ik}}{\partial x_k}, \tag{67}$$

$$\frac{\partial g_{ik}}{\partial t} = 2P_0(r)v_{ik} + v_j\nabla_j P_0(r)\delta_{ik}, \tag{68}$$

$$\pi_{ik} = 2\nu\rho v_{ik}, \quad v_{ik} = \frac{1}{2}\left(\frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i}\right). \tag{69}$$

The energy balance equation is

$$\frac{\partial}{\partial t}\int\frac{\rho v^2}{2}d^3\mathcal{V} = -\int[g_{ik} + \pi_{ik}]v_{ik}d^3\mathcal{V}. \tag{70}$$

This equation is obtained by scalar multiplication of (67) with v_i and integration over the cloud volume, provided that surface stresses are negligible. The next step is to use a separable $\mathbf{r}$ and t representation for both

the kinematic characteristics of motion like the field of material displacements and the rate-of-strain tensor,

$$v_i(\mathbf{r}, t) = a_i(\mathbf{r})\dot{\alpha}(t), \quad v_{ik}(\mathbf{r}, t) = a_{ik}(\mathbf{r})\dot{\alpha}(t), \quad (71)$$

and for the strength characteristics of motion such as the stress tensors of gravity and viscosity,

$$g_{ik}(\mathbf{r}, t) = [2P_0(r)a_{ik}(\mathbf{r}) + a_j(\mathbf{r})\nabla_j P_0(r)\delta_{ik}]\alpha(t), \quad \pi_{ik} = 2\eta a_{ik}\dot{\alpha}(t). \quad (72)$$

Substituting (71) and (72) in (70), we obtain the equation of damped harmonic oscillations

$$M\ddot{\alpha}(t) + D\dot{\alpha}(t) + K\alpha(t) = 0 \quad (73)$$

in which the parameters of inertia M, stiffness K, and viscous friction D are given by the integrals

$$M = \int \rho a_i(\mathbf{r})a_i(\mathbf{r})d^3V,$$

$$D = \frac{1}{2}\int \rho\nu[\nabla_i a_k + \nabla_k a_i][\nabla_i a_k + \nabla_k a_i]d^3V, \quad (74)$$

$$K = \frac{1}{2}\int P_0[\nabla_i a_k + \nabla_k a_i][\nabla_i a_k + \nabla_k a_i]d^3V. \quad (75)$$

These equations show that the frequency and lifetime of quasistatic modes of hydrogravity can be computed using the above specified fields of spheroidal and toroidal instantaneous displacements. Also, it is noteworthy that the last expression for K is similar to the equation for the rigidity coefficient of a viscoelastic material whose oscillatory response is controlled by Hooke's restoring force. This again lead us to conclude that the gravitational stresses imparts to the gas–dust circumstellar material a portion of shear mechanical rigidity typical of viscoelastic soft matter. Deserving particular emphasis is the fact that the parameter of rigidity (75) computed for both the spheroidal hydrogravity mode (with field of displacement (36)) and toroidal hydrogravity mode (with fields of displacement (37)) is analytically identical with that given by Eqs. (40) and (46). Therefore, all physically significant results inferred in the body of this paper can be recovered within the approach outlined in this appendix.

REFERENCES

1. D. Pines, J. Shaham, and M. A. Ruderman, in *Physics of Dense Matter: IAU Symposium*, Ed. by C. J. Hansen (Reidel, Dordrecht, 1974), Vol. 53, p. 189.
2. V. Boriakoff, Astron. J. **208**, L43 (1976).
3. H. M. Van Horn, Astrophys. J. **236**, 899 (1980).
4. G. S. Bisnovatyi-Kogan, V. S. Imshennik, D. K. Nadyozhin, and V. M. Chechetkin, Astrophys. Space Sci. **35**, 23 (1975).
5. O. Blaes, R. Blandford, P. Goldreich, and P. Madau, Astrophys. J. **343**, 839 (1989).
6. P. M. McDermott, H. M. Van Horn, and C. J. Hansen, Astrophys. J. **325**, 725 (1988).
7. T. E. Strohmayer, Astrophys. J. **372**, 573 (1991).
8. S. I. Bastrukov, V. V. Papoyan, D. V. Podgainy, and F. Weber, Nucl. Phys. B (Proc. Suppl.) **80**, 242 (2000).
9. S. I. Bastrukov, I. V. Molodtsova, D. V. Podgainy, *et al.*, Phys. Part. Nucl. **30**, 436 (1999).
10. S. I. Bastrukov, D. V. Podgainy, J. Yang, and F. Weber, Zh. Éksp. Teor. Fiz. **122**, 915 (2002) [JETP **95**, 789 (2002)].
11. S. Bastrukov, V. Papoyan, D. Podgainy, and J. Yang, Astrophysics **45**, 232 (2002).
12. H. Lamb, *Hydrodynamics*, 6th ed. (Cambridge Univ. Press, Cambridge, 1932; Gostekhizdat, Moscow, 1947).
13. M. J. Lighthill, *Waves in Fluids* (Cambridge Univ. Press, Cambridge, 1978; Mir, Moscow, 1981).
14. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics*, 3rd ed. (Nauka, Moscow, 1986; Pergamon, New York, 1959).
15. K. R. Lang, *Astrophysical Formulas* (Springer, Berlin, 1980).
16. S. I. Bastrukov, Int. J. Mod. Phys. D **5**, 45 (1996).
17. V. A. Fock, *The Theory of Space, Time, and Gravity* (Pergamon, New York, 1959).
18. S. Chandrasekhar, Astrophys. J. **158**, 45 (1969).
19. C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1970; Mir, Moscow, 1977).
20. F. H. Shu, *Physics of Astrophysics* (Univ. Sci. Books, Mill Valley, 1992).
21. L. D. Landau, E. M. Lifshitz, and L. P. Pitaevskii, *Electrodynamics of Continuous Media* (Butterworth-Heinemann, Oxford, 1995).
22. L. Mestel, *Stellar Magnetism* (Clarendon Press, Oxford, 1999).
23. G. S. Bisnovatyi-Kogan, *Stellar Physics* (Springer, Berlin, 2002).
24. S. I. Bastrukov, Mod. Phys. Lett. A **8**, 711 (1993).
25. S. I. Bastrukov, Phys. Rev. E **53**, 1917 (1996).
26. S. I. Bastrukov, J. Yang, D. V. Podgainy, and F. Weber, J. Phys. G **29**, 683 (2003).
27. A. N. Timokhin, G. S. Bisnovatyi-Kogan, and H. Spriut, Mon. Not. R. Astron. Soc. **316**, 734 (2000).
28. T. E. Strohmayer, J. M. Cordes, and H. M. Van Horn, Astrophys. J. **389**, 685 (1992).
29. J. Christensen-Dalsgaard, Rev. Mod. Phys. **74**, 1073 (2002).
30. A. B. Severny, V. A. Kotov, and T. T. Tsap, Nature **259**, 87 (1976).
31. J. R. Brookes, G. R. Isaak, and H. B. van der Raay, Nature **259**, 92 (1976).
32. A. B. Severny, V. A. Kotov, and T. T. Tsap, in *Proceedings of IAU Symposium on Advances in Helio- and Asteroseismology* (1988), Vol. 123, p. 33.
33. Y. P. Elsworth, S. M. Jefferis, C. P. McLeod, *et al.*, Astrophys. J. **338**, 557 (1989).
34. V. A. Kotov, V. I. Khaneychuk, and T. T. Tsap, Kinematika Fiz. Nebesnykh Tel **16**, 49 (2000).

Journal of Experimental and Theoretical Physics, Vol. 99, No. 3, 2004, pp. 460–465.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 3, 2004, pp. 527–532.

GRAVITATION, ASTROPHYSICS

Spectrum of a Quantized Black Hole, Correspondence Principle, and Holographic Bound¶

I. B. Khriplovich

Budker Institute of Nuclear Physics, Novosibirsk, 630090 Russia
Novosibirsk University, Novosibirsk, 630090 Russia
e-mail: khriplovich@inp.nsk.su

Received April 19, 2004

Abstract—An equidistant spectrum of the horizon area of a quantized black hole does not follow from the correspondence principle or from general statistical arguments. On the other hand, such a spectrum obtained in loop quantum gravity (LQG) either does not comply with the holographic bound or requires a special choice of the Barbero–Immirzi parameter for the horizon surface, distinct from its value for other quantized surfaces. The problem of distinguishability of the edges in LQG is discussed, with the following conclusion: Only under the assumption of partial distinguishability of the edges can the microcanonical entropy of a black hole be made both proportional to the horizon area and satisfying the holographic bound. © *2004 MAIK "Nauka/Interperiodica".*

1. The idea of quantizing the horizon area of black holes was put forward many years ago by Bekenstein in the pioneering article [1]. It was based on an intriguing observation made by Christodoulou and Ruffini [2, 3]: the horizon area of a nonextremal black hole behaves, in a sense, like an adiabatic invariant. Of course, the quantization of an adiabatic invariant is perfectly natural, in accordance with the correspondence principle.

One more conjecture made in [1] is that the spectrum of a quantized horizon area is equidistant. The argument therein was that a periodic system is quantized by equating its adiabatic invariant to $2\pi\hbar n$, $n = 0, 1, 2, \ldots$.

Later, it was pointed out by Bekenstein [4] that the classical adiabatic invariance does not by itself guarantee the equidistance of the spectrum, at least because any function of an adiabatic invariant is itself an adiabatic invariant. However, up to now, articles on the subject abound in assertions that the form

$$A = \beta l_p^2 n, \quad n = 1, 2, \ldots, \tag{1}$$

for the horizon area spectrum[1] is dictated by the respectable correspondence principle. The list of these references is too long to be presented here.

We consider an instructive example of the situation where a nonequidistant spectrum arises in spite of the classical adiabatic invariance. We start with a classical spherical top of an angular momentum **J**. Of course, the z projection J_z of **J** is an adiabatic invariant. If the z axis is chosen along **J**, the value of J_z is J, or $\hbar j$ in the quantum case. The classical angular momentum squared J^2 is also an adiabatic invariant, with the eigenvalues $\hbar^2 j(j+1)$ when quantized. We now try to use the operator $\hat{J}^2$ for the area quantization in quite natural units of l_p^2. For the horizon area A to be finite in the classical limit, the power of the quantum number j in the result for $j \gg 1$ should be the same as that of $\hbar$ in l_p^2 [5]. With $l_p^2 \sim \hbar$, we thus arrive at

$$A \sim l_p^2 \sqrt{j(j+1)}.$$

Because

$$\sqrt{j(j+1)} \longrightarrow j + 1/2 \text{ for } j \gg 1,$$

we have returned to the equidistant spectrum in the classical limit. However, the equidistant spectrum can be avoided as follows. We assume that the horizon area consists of sites with area of the order l_p^2, and to each site i we ascribe its own quantum number j_i and the contribution $\sqrt{j_i(j_i+1)}$ to the area. Then the above formula changes to

$$A \sim l_p^2 \sum_i \sqrt{j_i(j_i+1)} \tag{2}$$

(in fact, this formula for a quantized area arises as a special case in loop quantum gravity, see below). Of

¶ This article was submitted by the author in English.

[1] Here and below, $l_p^2 = \hbar k/c^3$ is the Planck length squared, $l_p = 1.6 \times 10^{-33}$ cm, and k is the Newton gravitational constant; β is here some numerical factor.

course, to retain a finite classical limit for A, we should require that

$$\sum_i \sqrt{j_i(j_i+1)} \gg 1.$$

However, any of the j_i can be comparable to unity. Therefore, in spite of the adiabatic invariance of A, its quantum spectrum (2) is not equidistant, although it is, of course, discrete.

One more, quite popular argument in favor of equidistant spectrum (1) is as follows [4, 6, 7]. On the one hand, the entropy S of a horizon is related to its area A by the Bekenstein–Hawking formula

$$A = 4l_p^2 S. \tag{3}$$

On the other hand, the entropy is nothing else but $\ln g(n)$, where the statistical weight $g(n)$ of any quantum state n is an integer. In [4, 6, 7], the requirement of integer $g(n)$ is taken literally and results after simple reasoning not only in equidistant spectrum (1), but also in the following allowed values for the numerical factor β in this spectrum:

$$\beta = 4\ln k, \quad k = 2, 3, \ldots.$$

We can imagine, however, that with some model for S, $g(n)$ is given by a noninteger value of $K + \delta$, $0 < \delta < 1$, instead of an integer value K. Then, the entropy is

$$S = \ln(K+\delta) = \ln K + \delta/K.$$

Now, the typical value of the black hole's entropy

$$S = \ln K = \frac{A}{4l_p^2}$$

is huge, roughly 10^{76}. Therefore, the correction δ/K is absolutely negligible compared to $S = \ln K$. Moreover, it is far below any conceivable accuracy in a description of entropy and can therefore be safely omitted and forgotten. As usual for macroscopic objects, the fact that the statistical weight is an integer has no consequences for the entropy.

Thus, contrary to popular belief, the equidistance of the spectrum for the horizon area does not follow from the correspondence principle and/or from general statistical arguments.

2. This does not mean, however, that any model leading to an equidistant spectrum for the quantized horizon area should be automatically rejected. A quite simple and elegant version of such a model, so-called "it from bit," was formulated for a Schwarzschild black hole by Wheeler [8]. The assumption is that the horizon surface consists of ν patches, each of them supplied with an "angular momentum" quantum number j with two possible projections $\pm 1/2$. The total number K of degenerate quantum states of this system is

$$K = 2^\nu. \tag{4}$$

Then the entropy of the black hole is

$$S_{1/2} = \ln K = \nu \ln 2. \tag{5}$$

Finally, with Bekenstein–Hawking relation (3), one obtains the following equidistant formula for the area spectrum:

$$A_{1/2} = 4\ln 2 l_p^2 \nu. \tag{6}$$

This model of a quantized Schwarzschild black hole by itself looks flawless.

This result was later derived in [9] in the framework of LQG [10–14]. We discuss below whether the "it from bit" picture, if considered as a special case of the area quantization in LQG, can be reconciled with the holographic bound [15–17].

More generally, a quantized surface in LQG is described as follows. One ascribes a set of punctures to the surface. Each puncture is supplied with two integer or half-integer angular momenta j^u and j^d,

$$j^u, j^d = 1/2, 1, 3/2, \ldots. \tag{7}$$

j^u and j^d are related to edges directed up and down the normal to the surface, respectively, and add up to an angular momentum j^{ud},

$$\mathbf{j}^{ud} = \mathbf{j}^u + \mathbf{j}^d, \quad \left|j^u - j^d\right| \le j^{ud} \le j^u + j^d. \tag{8}$$

The area of the surface is

$$A = \beta l_p^2 \sum_i \sqrt{2j_i^u(j_i^u+1) + 2j_i^d(j_i^d+1) - j_i^{ud}(j_i^{ud}+1)}. \tag{9}$$

The overall numerical factor β in (9) cannot be determined without an additional physical input. This ambiguity originates from a free (so-called Barbero–Immirzi) parameter [18, 19] that corresponds to a family of inequivalent quantum theories, all of them being viable without such an input.

The result (6) was obtained in [9] under the additional condition that the gravitational field on the horizon is described by the $U(1)$ Chern–Simons theory. Formula (6) is a special case of general formula (9) with all j^d vanishing and all j^u equal to 1/2 (or vice versa). As regards the overall factor β, its value here is[2]

$$\beta = \frac{8\ln 2}{\sqrt{3}}. \tag{10}$$

We now turn to the holographic bound [15–17]. According to it, the entropy S of any spherically sym-

[2] The common convention for the numerical factor in formula (9) is $8\pi\beta$; with it, the parameter β is smaller than ours by the factor of 8π.

metric system confined inside a sphere of area A is bounded as

$$S \le \frac{A}{4l_p^2}, \tag{11}$$

with the equality attained only for a black hole.

A simple intuitive argument confirming this bound is as follows [17]. We consider the discussed system collapsing into a black hole. During the collapse, the entropy increases from S to S_{bh} and the resulting horizon area A_{bh} is certainly smaller than the initial confining area A. Now, with Bekenstein–Hawking relation (3) for a black hole taken into account, we arrive, via the obvious chain of (in)equalities

$$S \le S_{bh} = \frac{A_{bh}}{4l_p^2} \le \frac{A}{4l_p^2},$$

at the discussed bound (11).

The result (11) can be formulated in a different manner. Among the spherical surfaces of a given area, the surface of a black hole horizon has the largest entropy.

On the other hand, it is only natural that the entropy of an eternal black hole in equilibrium is maximum. This was used by Vaz and Witten [20] in a model of a quantum black hole originating from a dust collapse. The idea was then employed by us [21, 22] in the problem of quantizing the horizon of a black hole in LQG. In particular, the coefficient β was calculated in [22] in the case where the area of a black hole horizon is given by the general formula (9) of LQG, as well as under some more special assumptions on the values of j^u, j^d, and j^{ud}. Moreover, it was demonstrated in [22] for a rather general class of horizon quantization schemes that the maximum entropy of a quantized surface is proportional to its area.

We sketch the proof of this result (for more technical details, see [22]). Here and below, we consider the microcanonical entropy S of a surface (although with fixed area instead of fixed energy). It is defined as the logarithm of the number of states of this surface with a fixed area A, i.e., with a fixed sum

$$N = \sum_i \sqrt{2j_i^u(j_i^u+1) + 2j_i^d(j_i^d+1) - j_i^{ud}(j_i^{ud}+1)}. \tag{12}$$

Let ν_{im} be the number of punctures with a given set of momenta j_i^u, j_i^d, j_i^{ud}, and a given projection m of j_i^{ud}. The total number of punctures is

$$\nu = \sum_{i,m} \nu_{im}.$$

We assume that the edges with the same set of the quantum numbers i and m (i.e., with the same j_i^u, j_i^d, j_i^{ud}, and m) are indistinguishable, and therefore interchanging them does not result in new states. All other permutations, those among the edges with differing i and m, do create new states, and hence such edges, with differing i, m, are distinguishable,

We note that the "it from bit" values (4) and (5) for the number of states and entropy also follow from this assumption. Indeed, let ν be the total number of patches with $j = 1/2$ and let ν_+ and $\nu_- = \nu - \nu_+$ patches have the respective projections +1/2 and –1/2. Then the number of the corresponding states is obviously given by

$$\frac{\nu!}{\nu_+!(\nu-\nu_+)!},$$

and the total number of states is

$$K = \sum_{\nu_+=0}^{\nu} \frac{\nu!}{\nu_+!(\nu-\nu_+)!} = 2^\nu,$$

in agreement with (4).

Thus, the entropy is

$$S = \ln\left[\nu! \prod_{im} \frac{1}{\nu_{im}!}\right]. \tag{13}$$

The structure of expressions (9) and (13) is so different that the entropy certainly cannot be proportional to the area in the general case. However, this is the case for the maximum entropy in the classical limit.

For combinatorial reasons, it is natural to expect that the absolute maximum of entropy is reached when all values of quantum numbers $j_i^{u,d,ud}$ are present. We also assume that in the classical limit, the typical values of puncture numbers ν_{im} are large. Then, with the Stirling formula for factorials, expression (13) becomes

$$S = \left(\sum_{im} \nu_{im}\right) \ln\left(\sum_{i'm'} \nu_{i'm'}\right) - \sum_{im} (\nu_{im} \ln \nu_{im}). \tag{14}$$

We seek the extremum of expression (14) under the condition

$$N = \sum_i \nu_{im} r_i = \text{const}, \tag{15}$$

where each partial contribution

$$r_i = \sqrt{2j_i^u(j_i^u+1) + 2j_i^d(j_i^d+1) - j_i^{ud}(j_i^{ud}+1)}$$

is independent of m. The problem reduces to solving the system of equations

$$\ln\left(\sum_{i',m'} \nu_{i'm'}\right) - \ln \nu_{im} = \mu r_i, \tag{16}$$

or

$$\nu_{im} = e^{-\mu r_i}\sum_{i',m'}\nu_{i'm'} = \nu e^{-\mu r_i}. \quad (17)$$

Here, μ is the Lagrange multiplier for constraint (15). Summing expressions (17) over i and m, we arrive at the equation for μ,

$$\sum_{i,m} e^{-\mu r_i} = \sum_i g_i e^{-\mu r_i} = 1; \quad (18)$$

the statistical weight

$$g_i = 2j_i^{ud} + 1$$

of a puncture arises here because r_i are independent of m. On the other hand, multiplying Eq, (16) by ν_{im} and summing over i and m, with constraint (15), we arrive at the following result for the maximum entropy for a given value of the sum N, or the black hole area A:

$$S_{\max} = \mu N = \frac{\mu}{\beta l_p^2}A. \quad (19)$$

One more curious feature of the picture obtained is worth noting: it gives a sort of Boltzmann distribution for occupation numbers (17). In this distribution, the partial contributions r_i to the area are analogs of energies and the Lagrange multiplier μ corresponds (up to a factor) to inverse temperature.

It should be emphasized that relation (19) is true not only in LQG, but applies to a more general class of approaches to quantization of surfaces. The following assumption is necessary here: the surface should consist of patches of different sorts, such that there are ν_{im} patches of each sort i, m, with a generalized effective quantum number r_i and a statistical weight g_i. Equally necessary is the above assumption on the distinguishability of the patches.

Thus, the maximum entropy of a surface is proportional to its area in the classical limit. This proportionality certainly occurs for a classical black hole. This is one more strong argument in favor of the assumption that the black hole entropy is maximum.

We now return to the result in [9]. If we assume that the value (10) of the parameter β is universal (i.e., it is not special to black holes, but refers to any quantized spherical surface), then the value in (5) is not the maximum one in LQG for a surface of area (6). This looks quite natural: with the transition from the unique choice made in [9],

$$j^{u(d)} = 1/2, \quad j^{d(u)} = 0,$$

to a more extended and richer one, the number of degenerate quantum states should, generally speaking, increase. Together with this number, its logarithm, which is the entropy of a quantized surface, increases as well.

We begin the proof of the above statement by rewriting formula (5) as

$$S_{1/2} = \ln 2\sqrt{\frac{4}{3}}N = 0.80N, \quad N = \sqrt{\frac{3}{4}}\nu. \quad (20)$$

From now on, we consider this value of N fixed. We start with a relatively simple example where

$$j^{d(u)} = 0,$$

and hence the general formula (9) for a surface area reduces to

$$A = \beta l_p^2\sum_i\sqrt{j_i(j_i+1)} = \beta l_p^2\sum_{j=1/2}^{\infty}\sqrt{j(j+1)}\nu_j, \quad j = j^{u(d)} \quad (21)$$

(and coincides with our naive model (2)). We find the maximum entropy of such a surface for the fixed value of

$$N = \sum_{j=1/2}^{\infty}\sqrt{j(j+1)}\nu_j, \quad (22)$$

which should be equal to the "it from bit" one, $\nu\sqrt{3/4}$. Here, the statistical weight of a puncture with quantum number j is

$$g_j = 2j + 1,$$

and Eq. (18) can be rewritten as

$$\sum_{p=1}^{\infty}(p+1)z^{\sqrt{p(p+2)}} = 1, \quad p = 2j, \quad z = e^{-\mu/2}. \quad (23)$$

Its solution is

$$\mu = -2\ln z = 1.722$$

(see [22]), and the maximum entropy

$$S_{\max,1} = 1.72N \quad (24)$$

then exceeds the result in (20).

As expected, in the general case, with N given by formula (12) with all the values of j_i^u, j_i^d, j_i^{ud} allowed and

$$g_i = 2j_i^{ud} + 1,$$

the maximum entropy is even larger [22],

$$S_{\max} = 3.12N. \quad (25)$$

Thus, the conflict is obvious between the holographic bound and the result (20) found within the LQG approach in [9].

One might try to avoid the conflict by assuming that value (10) of the Barbero–Immirzi parameter β is special for black holes only, while for other quantized sur-

faces, β is smaller. However, such a way out would be unattractive and unnatural.

3. We now return to the essential assumption made in the previous section: the edges with the same set of the quantum numbers i, m are identical, and the edges with differing i, m are distinguishable. In principle, one might try to modify this assumption of partial distinguishability of edges in two opposite ways.

One possibility that might look quite appealing is that of complete indistinguishability of edges. It means that no permutation of any edges results in new states. To simplify the discussion, we consider to expression (21) for the horizon area instead of the most general one (9).

Then, the total number of angular momentum states created by

$$\nu_j = \sum_m \nu_{jm}$$

indistinguishable edges of a given j with all $2j + 1$ projections allowed, from $-j$ to j, is[3]

$$K_j = \frac{(\nu_j + 2j)!}{\nu_j!(2j)!}. \quad (26)$$

Those partial contributions

$$s_j = \ln K_j$$

to the black hole's entropy

$$S = \sum_j s_j,$$

which can potentially dominate the numerically large entropy, may correspond to three cases: $j \ll \nu_j$, $j \gg \nu_j$, and $j \sim \nu_j \gg 1$. These contributions are as follows:

$$j \ll \nu_j, \quad s_j \approx 2j\ln\nu_j;$$

$$j \gg \nu_j, \quad s_j \approx \nu_j \ln j;$$

$$j \sim \nu_j \gg 1, \quad s_j \sim 4j\ln 2.$$

In all three cases, the partial contributions to the entropy S are much smaller parametrically than the corresponding contributions

$$a_j \sim j\nu_j$$

to the area

$$A = \sum_j a_j.$$

Therefore, $S \ll A$ in all these cases, and hence, with indistinguishable edges of the same j, one cannot make the entropy of a black hole proportional to its area. This was pointed out earlier in [23, 24].

We now consider the last conceivable option, that of completely distinguishable edges. In this case, the total number of states is just $K = \nu!$, instead of (13), with the microcanonical entropy

$$S = \nu\ln\nu.$$

In principle, this entropy can be made proportional to the black hole's area A. The model (which does not look natural, however) could be as follows. We choose a large quantum number $J \gg 1$ and assume that the horizon area A is saturated by the edges with j in the interval $J < j < 2J$ and with "occupation numbers" $\nu_j \sim \ln J$. Then the estimates for both S and A are of the order of $J\ln J$, and the proportionality between the entropy and the area can be attained.

However, although the entropy can be proportional to the area under the assumption of complete distinguishability, the maximum entropy for a given area is much larger than the area itself. Obviously, the maximum entropy for fixed

$$A \sim \sum_j \sqrt{j(j+1)}\nu_j$$

is here attained with all j's being as small as possible, e.g., 1/2 or 1. In the classical limit $\nu \gg 1$, the entropy of a black hole then grows faster than its area, $A \sim \nu$, while

$$S = \nu\ln\nu \sim A\ln A.$$

Thus, the assumption of complete distinguishability is in conflict with the holographic bound and therefore should be discarded.

There is no disagreement between this conclusion and that made in [23, 25, 26]: what is called complete distinguishability therein corresponds to our partial distinguishability.

ACKNOWLEDGMENTS

I am grateful to V.F. Dmitriev, V.M. Khatsymovsky, and G.Yu. Ruban for discussions. Research was supported in part by the Russian Foundation for Basic Research (project no. 03-02-17612).

REFERENCES

1. J. D. Bekenstein, Lett. Nuovo Cimento **11**, 467 (1974).
2. D. Christodoulu, Phys. Rev. Lett. **25**, 1596 (1970).
3. D. Christodoulu and R. Ruffini, Phys. Rev. D **4**, 3552 (1971).
4. J. D. Bekenstein, in *Proceedings of Eighth Marcel Grossmann Meeting on General Relativity*, Ed. by Tsvi Piran and Remo Ruffini (World Sci., Singapore, 1999), p. 92; gr-qc/9710076.
5. I. B. Khriplovich, Phys. Lett. B **431**, 19 (1998); gr-qc/9804004.

[3] Perhaps, the simplest derivation of this formula is as follows. Effectively, we here seek the number of ways of distributing ν_j identical balls into $2j + 1$ boxes. Then, the line of reasoning presented in [27, Section 54] results in formula (26). I am grateful to V.F. Dmitriev for bringing to my attention that formula (26) can be derived in this simple way.

6. J. D. Bekenstein and V. F. Mukhanov, Phys. Lett. B **360**, 7 (1995); gr-qc/9505012.
7. J. D. Bekenstein and V. F. Mukhanov, in *Proceedings of Sixth Moscow Quantum Gravity Seminar*, Ed. by V. A. Berezin, V. A. Rubakov, and D. V. Semicoz (World Sci., Singapore, 1997).
8. J. A. Wheeler, in *Sakharov Memorial Lectures in Physics: Proceedings of the First International Sakharov Conference on Physics*, Ed. by L. V. Keldysh and V. Ya. Feinberg (Moscow, 1991), Vol. 2, p. 751.
9. A. Ashtekar, J. Baez, A. Corichi, and K. Krasnov, Phys. Rev. Lett. **80**, 904 (1998); gr-qc/9710007.
10. C. Rovelli and L. Smolin, Nucl. Phys. B **442**, 93 (1995); Nucl. Phys. B **456**, 753(E) (1995); gr-qc/9411005.
11. A. Ashtekar and J. Lewandowski, Class. Quantum Grav. **14**, 55 (1997); gr-qc/9602046.
12. R. Loll, Phys. Rev. Lett. **75**, 3048 (1995); gr-qc/9506014; Nucl. Phys. B **460**, 143 (1996); gr-qc/9511030.
13. R. De Pietri and C. Rovelli, Phys. Rev. D **54**, 2664 (1996); gr-qc/9602023.
14. S. Frittelli, L. Lehner, and C. Rovelli, Class. Quantum Grav. **13**, 2921 (1996); gr-qc/9608043.
15. J. D. Bekenstein, Phys. Rev. D **23**, 287 (1981).
16. G. 't Hooft, in *Salam Festschrift* (World Sci., Singapore, 1993); gr-qc/9310026.
17. L. Susskind, J. Math. Phys. **36**, 6377 (1995); gr-qc/9710007.
18. J. F. Barbero, Phys. Rev. D **51**, 5507 (1995); gr-qc/9410014.
19. G. Immirzi, Class. Quantum Grav. **14**, L177 (1997); gr-qc/9701052.
20. C. Vaz and L. Witten, Phys. Rev. D **64**, 084005 (2001); gr-qc/0104017.
21. I. B. Khriplovich, Phys. Lett. B **537**, 125 (2002); gr-qc/0109092.
22. R. V. Korkin and I. B. Khriplovich, Zh. Éksp. Teor. Fiz. **122**, 5 (2002) [JETP **95**, 1 (2002)]; gr-qc/0112074.
23. A. Alekseev, A. P. Polychronakos, and M. Smedback, Phys. Lett. B **574**, 296 (2003); hep-th/0004036.
24. I. B. Khriplovich, in *Proceedings of St. Petersburg School of Physics* (St. Petersburg, 2002); gr-qc/0210108.
25. A. P. Polychronakos, Phys. Rev. D **69**, 044010 (2004); hep-th/0304135.
26. G. Gour and V. Suneeta, Class. Quantum Grav. **21**, 3405 (2004); gr-qc/0401110.
27. L. D. Landau and E. M. Lifshitz, *Statistical Physics*, 2nd ed. (Nauka, Moscow, 1964; Pergamon, London, 1958).

Journal of Experimental and Theoretical Physics, Vol. 99, No. 3, 2004, pp. 466–473.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 3, 2004, pp. 533–541.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Excitation of Strong Wakefields by Intense Neutrino Bursts in a Magnetized Electron–Positron Plasma¶

A. Serbeto[a], L. A. Rios[a], J. T. Mendonça[b], P. K. Shukla[c], and R. Bingham[d]

[a]*Instituto de Física, Universidade Federal Fluminense 24210-340, Rio de Janeiro, Brazil*
[b]*GoLP/Instituto Superior Técnico, Universidade Técnica de Lisboa 1049-001, Lisbon, Portugal*
[c]*Institut für Theoretische Physik IV, Ruhr-Universität Bochum D-44780, Bochum, Germany*
[d]*Rutherford Appleton Laboratory, Chilton, Didcot Oxon OX11 0QX, England*

Received April 12, 2004

Abstract—A classical fluid description is used to investigate nonlinear interactions between an electron-type neutrino burst and a collisionless magnetized electron–positron plasma. It is found that the symmetry between the electron and positron dynamics is broken due to the presence of intense neutrino bursts. The latter can excite strong upper-hybrid wakefields, which can produce unlimited acceleration of pairs across the external magnetic field direction via a surfatron mechanism. Implications of our results to the production of high-energy electrons and positrons in astrophysical environments are discussed. *© 2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is well established that processes involving electron–positron plasmas are of significant importance in a variety of astrophysical phenomena [1–4] that presumably took place in the early Universe, during the period $10^{-4} < t < 1$ s after the Big Bang, and in ultrarelativistic pair plasma jets. Electron–positron plasmas are also found in active galactic nuclei [5, 6] and in the pulsar magnetosphere [7, 8]. Another important phenomenon is the γ-ray burst [9, 10], where many of its characteristics are explained via the relativistic expansion of the electron–positron plasma ("the fireball model") [10]. It is well known that such a plasma has a very peculiar nature compared with the traditional electron–ion plasma, because electrons and positrons have the same mass and represent a symmetry due to their opposite electric charges. These unique properties are responsible for many linear and nonlinear wave phenomena that are different from those occurring in an electron–ion plasma. This difference is due to the electron–ion mass ratio, which gives rise to different timescales associated with the electron and ion dynamics in the plasma. A systematic study (see, e.g., [11] and references therein) of nonlinear interactions between intense electromagnetic waves and relativistic electron–positron plasmas has been carried out, especially in the pulsar pair plasma environment for understanding the origin of the pulsar radio emission.

Other very important material elements in astrophysical settings are neutrinos. They are produced by the core of stars and in very high-explosive astrophysical situations such as those in supernova explosions and in ultrarelativistic pair plasma jets producing γ-ray bursts. Recently, many authors [12–15] have proposed that the shock expansion mechanism in the supernova could be due to the energy-momentum transfer from the neutrino bursts to the magnetized plasma cloud that surrounds the core of stars. Here, we are interested in nonlinear interactions of intense neutrino bursts with a relativistic magnetized pair plasma. It is well known [16–19] that neutrinos propagating through the plasma can acquire an effective (induced) electric charge due to charged and neutral currents associated with the weak nuclear force causing exchange of $W^{\pm}$ and Z^0 bosons. There also appears a nonlinear coupling between the neutrinos and the plasma via the weak Fermi nuclear interaction and the ponderomotive force of the neutrino beam [20–25].

In this paper, we study the generation of large-amplitude upper-hybrid waves in a magnetized electron–positron plasma interacting with an electron-type neutrino burst. The neutrino dynamics can be considered semiclassical by assuming that the neutrino de Broglie wavelength is much shorter than the typical scalelength of the perturbation in the effective neutrino weak interaction potential. We ignore all quantum-mechanical effects (e.g., the neutrino magnetic moment) caused by external magnetic fields, because

$$\frac{B_0}{B_{QM}} \equiv \frac{\hbar/\omega_{ce}}{m_e c^2} \ll 1,$$

where $\hbar$ is the Planck constant divided by 2π,

$$\omega_{ce} = \frac{eB_0}{m_e c}$$

is the electron gyrofrequency, e is the magnitude of the electron charge, m_e is the electron mass, c is the speed

¶This article was submitted by the authors in English.

of light in vacuum, and $B_{QM} \approx 4 \times 10^{13}$ Gs is the Landau–Schwinger critical field. Neutral current forward scattering of neutrinos off neutrinos in the background makes a contribution proportional to the density matrix, which has no effect on flavor evolution [26].

This paper is organized as follows. In Section 2, we present the governing equations for neutrinos and electrostatic waves. Section 3 contains an analytic description for the upper-hybrid wakefield in the presence of neutrino fluxes. Numerical results for the wakefield are presented in Section 4. Finally, Section 5 highlights our results and contains possible applications of our work to acceleration of pairs by large-amplitude upper-hybrid waves in astrophysical environments.

2. GOVERNING EQUATIONS

We consider an electron-type neutrino burst in a magnetized electron–positron plasma. The external magnetic field $B_0\hat{\mathbf{z}}$ is along the z direction. The dynamics of an ensemble of the neutrinos can be described by the equations [22]

$$\frac{\partial N_\nu}{\partial t} + \nabla \mathbf{J}_\nu = 0 \tag{1a}$$

and

$$\frac{\partial \mathbf{p}_\nu}{\partial t} + (\mathbf{v}_\nu \nabla)\mathbf{p}_\nu = -\frac{1}{N_\nu}\nabla P_\nu + \sum_\sigma G_{\sigma\nu}\left(\mathbf{E}_\sigma + \frac{\mathbf{v}_\nu}{c}\times \mathbf{B}_\sigma\right), \tag{1b}$$

which couple the neutrino density N_ν and the neutrino momentum $\mathbf{p}_\nu$. Here,

$$G_{\sigma\nu} = \sqrt{2}G_F[\delta_{\sigma e}\delta_{\nu\nu_e} + (I_\sigma - 2Q_\sigma \sin^2\theta_w)]$$

is the "bare" charge, with $G_{\sigma\nu} = -G_{\bar{\sigma}\nu}$, which allows the neutrinos to couple to the plasma fluid. Furthermore, σ denotes the electron (e^-) and positron (e^+) species of the plasma, G_F ($\approx 9 \times 10^{-30}$ eV cm^{-3}) is the Fermi weak-interaction coupling constant, θ_W is the Weinberg mixing angle ($\sin^2\theta_W \approx 0.23$), I_σ is the weak isotopic spin of the particle of the species σ (equal to –1/2 and 1/2 for the electrons and positrons, respectively), and $Q_\sigma = q_\sigma/e$ is the particle normalized electric charge. It should be noted that the first term in the neutrino "bare" charge is due to charged currents, which is valid only for electrons (positrons) and the electron-type neutrinos, and other terms come from the neutral currents and are valid for all particle species.

In Eq. (1b), $P_\nu = N_\nu T_\nu$ is the neutrino kinetic pressure and the second term in Eq. (1b) is the weak force, $\mathbf{F}_\nu$, on a single neutrino due to the plasma. Furthermore,

$$\mathbf{E}_\sigma = -\nabla N_\sigma - \frac{1}{c^2}\frac{\partial \mathbf{J}_\sigma}{\partial t},$$

$$\mathbf{B}_\sigma = c^{-1}\nabla \times \mathbf{J}_\sigma$$

are the effective electric and magnetic fields [22], respectively, and

$$\mathbf{J}_\nu = \mathbf{v}_\nu N_\nu, \quad \mathbf{J}_\sigma = \mathbf{v}_\sigma N_\sigma$$

are the neutrino and σ species currents, respectively. The linear momentum of the neutrino is

$$\mathbf{p}_\nu = \frac{\mathbf{v}_\nu}{c^2}E_\nu,$$

with E_ν being the neutrino total energy. The term $\partial \mathbf{J}_\sigma/\partial t$ is the neutrino-plasma (treated covariantly) analog of the electromagnetic-plasma energy transfer, as described in [22]. Furthermore, strong dc magnetic fields can create magnetic-field-aligned motion of the neutrinos and plasma particles.

The plasma particle dynamics is governed by the continuity and momentum equations, which are, respectively,

$$\frac{\partial N_\sigma}{\partial t} + \nabla \mathbf{J}_\sigma = 0 \tag{2a}$$

and

$$\frac{\partial \mathbf{P}_\sigma}{\partial t} + (\mathbf{v}_\sigma \nabla)\mathbf{P}_\sigma = q_\sigma \mathbf{E} + q_\sigma \frac{\mathbf{v}_\sigma}{c}\times \mathbf{B} + \sum_\nu G_{\sigma\nu}\left(\mathbf{E}_\nu + \frac{\mathbf{v}_\sigma}{c}\times \mathbf{B}_\nu\right), \tag{2b}$$

where

$$\mathbf{P}_\sigma = \gamma_\sigma m_\sigma \mathbf{v}_\sigma$$

is the momentum of the particle species σ (electrons and positrons),

$$\gamma_\sigma = \frac{1}{\sqrt{1 - v_\sigma^2/c^2}}$$

is the gamma factor, and $\mathbf{B} = B_0\hat{\mathbf{z}}$ is the external uniform magnetic field in the z direction. The right-hand side in Eq. (2b) is the total force acting on the plasma due to all types of the neutrinos, and

$$\mathbf{E}_\nu = -\nabla N_\nu - c^{-2}\frac{\partial \mathbf{J}_\nu}{\partial t},$$

$$\mathbf{B}_\nu = c^{-1}\nabla \times \mathbf{J}_\nu$$

are "weak-electromagnetic" fields. Furthermore, N_σ is the number density of the species σ. Because we focus

on the wakefield generation on timescales that are either comparable to or shorter than the electron plasma period, collisions between pairs do not play an essential role, for instance, in supernova and pair plasma jets [24, 25], where the plasma number density and the pair temperature are 10^{30} cm^{-3} and 10^5–10^6 eV, respectively.

To simplify our model, we consider only a cold electron-type neutrino streaming along the x direction with a velocity v_ν close to c, interacting nonlinearly with a collisionless cold magnetized electron–positron plasma. Hence, antineutrinos are neglected. It is well known that the interactions of electron neutrinos with the plasma do not change their local energy and density significantly. For instance, in type-II supernova explosions, only 1% of the neutrino energy [14] is supposed to be transferred to the plasma that surrounds the core of the star. Hence, without loss of generality, we can assume that the electron-type neutrino flux only transfers a very small part of its energy E_ν to the plasma and keeps its density N_ν nearly constant. Accordingly, the electron-type neutrino fluid dynamical equations can be rewritten as

$$\frac{\partial E_\nu}{\partial t} + c\frac{\partial E_\nu}{\partial x} \approx -\sqrt{2}G_F c\left(\frac{\partial}{\partial x}(N_{\bar e} - N_e) + \frac{1}{c^2}\frac{\partial}{\partial t}(J_{\bar e} - J_e)\right) \tag{3a}$$

and

$$\frac{\partial N_\nu}{\partial t} + c\frac{\partial N_\nu}{\partial x} \approx 0. \tag{3b}$$

With the longitudinal plasma waves propagating across the external magnetic field direction, with an associated electric field $\mathbf{E} = E\hat{\mathbf{x}}$ and wavenumber $\mathbf{k} = k\hat{\mathbf{x}}$, the electron–positron plasma fluid equations are

$$\frac{\partial N_e}{\partial t} + \frac{\partial J_e}{\partial x} = 0, \tag{4a}$$

$$\frac{\partial P_{ex}}{\partial t} + \frac{cP_{ex}}{\sqrt{1+P_e^2}}\frac{\partial P_{ex}}{\partial x} = -\frac{eE}{m_e c} - \frac{\omega_c P_{ey}}{\sqrt{1+P_e^2}} - \sqrt{2}G_F\left(\frac{\partial N_\nu}{\partial x} + \frac{1}{c^2}\frac{\partial J_\nu}{\partial t}\right), \tag{4b}$$

$$\frac{\partial P_{ey}}{\partial t} + \frac{cP_{ey}}{\sqrt{1+P_e^2}}\frac{\partial P_{ey}}{\partial x} = \frac{\omega_c P_{ex}}{\sqrt{1+P_e^2}}, \tag{4c}$$

and

$$\frac{\partial N_{\bar e}}{\partial t} + \frac{\partial J_{\bar e}}{\partial x} = 0, \tag{5a}$$

$$\frac{\partial P_{\bar e x}}{\partial t} + \frac{cP_{\bar e x}}{\sqrt{1+P_{\bar e}^2}}\frac{\partial P_{\bar e x}}{\partial x} = \frac{eE}{m_e c} + \frac{\omega_c P_{\bar e x}}{\sqrt{1+P_{\bar e}^2}} + \sqrt{2}G_F\left(\frac{\partial N_\nu}{\partial x} + \frac{1}{c^2}\frac{\partial J_\nu}{\partial t}\right), \tag{5b}$$

$$\frac{\partial P_{\bar e y}}{\partial t} + \frac{cP_{\bar e y}}{\sqrt{1+P_e^2}}\frac{\partial P_{\bar e y}}{\partial x} = -\frac{\omega_c P_{\bar e x}}{\sqrt{1+P_{\bar e}^2}} \tag{5c}$$

for the electron and positron plasma species, with P_{ex}, $P_{\bar e x}$, P_{ey}, and $P_{\bar e y}$ being the x and y components of the electron and positron momenta, respectively.

Equations (3), (4), and (5) form a set for studying the generation of large-amplitude plasma waves. To eliminate the electron and positron fluid currents in Eq. (3a), we supplement our system of equations with

$$\mathbf{J}_{\bar e} - \mathbf{J}_e = \frac{1}{4\pi e}\frac{\partial \mathbf{E}}{\partial t}, \tag{6}$$

because we consider the generation of longitudinal (electrostatic) waves.

3. UPPER-HYBRID WAKEFIELD

It is convenient to introduce a new independent variable $\xi = (x - v_\phi t)$, where v_ϕ is the plasma wave phase speed. Hence, Eq. (3a) can be rewritten as

$$\frac{dE}{d\xi} = 4\pi e\frac{c^2}{v_\phi^2}\left((N_{\bar e} - N_e) + \frac{1-\beta_\phi}{\sqrt{2}G_F}(E_0 - E_\nu(\xi))\right), \tag{7}$$

where E_0 is the neutrino initial energy. Using the definition $E = -d\Phi/d\xi$, where Φ is the electric potential associated with the wakefield, we obtain from Eq. (7)

$$\frac{d^2\Psi}{d\xi^2} = -\frac{\omega_p^2}{v_\phi^2}\left(\frac{N_{\bar e}}{N_0} - \frac{N_e}{N_0} + S_\nu\right), \tag{8}$$

where

$$\Psi = \frac{e\Phi}{m_e c^2}$$

is the normalized plasma potential,

$$\beta_\phi = \frac{v_\phi}{c}$$

is the normalized phase speed,

$$\omega_p^2 = \frac{4\pi e^2 N_0}{m_e}$$

is the squared electron plasma frequency, and N_0 is the equilibrium electron (positron) number density. Furthermore,

$$S_\nu = \frac{E_0(1-\beta_\phi)}{\sqrt{2}G_F N_0}\frac{\Delta E_\nu}{E_0}$$

represents the neutrino-driven term, with

$$\frac{\Delta E_\nu}{E_0} = \frac{\Delta\omega_\nu}{\omega_\nu}$$

being considered the amount of neutrino energy transferred to the plasma. Here, $\Delta\omega_\nu$ is the spectral width of the neutrino spectrum. Assuming that the magnitude

$$\frac{\Delta E_\nu}{E_0} = \frac{\Delta\omega_\nu}{\omega_\nu} \ll 1$$

(actually, from the observational fact in the supernova SN1987A, the "visible" energy of the supernova is a very small part of the neutrino energy [14]), we can consider the neutrino flux as an external action into the plasma such that the amount of the neutrino energy deposited in the plasma can be taken as a constant input in Eq. (8). We note that the latter is the Poisson equation written in a moving frame, where the total charge density includes the neutrino effective charge density represented by the term S_ν. Transforming Eqs. (4) and (5) and substituting the electron (positron) density response into Eq. (8), we obtain

$$\frac{d^2\Psi}{d\chi^2} = \frac{\Gamma_e\sqrt{1+P_e^2}}{P_{ex}-\beta_\phi\sqrt{1+P_e^2}} - \frac{\Gamma_{\bar e}\sqrt{1+P_{\bar e}^2}}{P_{\bar e x}-\beta_\phi\sqrt{1+P_{\bar e}^2}} - S_\nu, \quad (9)$$

$$\frac{dP_{ex}}{d\chi} = \frac{\sqrt{1+P_e^2}}{P_{ex}-\beta_\phi\sqrt{1+P_e^2}}\left(\frac{d\Psi}{d\chi} - \beta_\phi\Omega_c\frac{P_{ey}}{\sqrt{1+P_e^2}}\right), \quad (10)$$

$$\frac{dP_{ey}}{d\chi} = \beta_\phi\Omega_c\frac{P_{ex}}{(P_{ex}-\beta_\phi\sqrt{1+P_e^2})}, \quad (11)$$

$$\frac{dP_{\bar e x}}{d\chi} = \frac{\sqrt{1+P_{\bar e}^2}}{P_{\bar e x}-\beta_\phi\sqrt{1+P_{\bar e}^2}}\left(-\frac{d\Psi}{d\chi} + \beta_\phi\Omega_c\frac{P_{\bar e y}}{\sqrt{1+P_{\bar e}^2}}\right), \quad (12)$$

and

$$\frac{dP_{\bar e y}}{d\chi} = -\beta_\phi\Omega_c\frac{P_{\bar e x}}{(P_{\bar e x}-\beta_\phi\sqrt{1+P_e^2})}, \quad (13)$$

where

$$P_{e,\bar e} = p_{e,\bar e}/m_e c$$

is the normalized electron (positron) momentum,

$$\chi = \omega_p/v_\phi(x - v_\phi t)$$

is the normalized distance (phase), and

$$\Omega_c = \omega_c/\omega_p$$

is the normalized electron (positron) gyrofrequency. Here,

$$\Gamma_{e,\bar e} = \frac{P_0 - \beta_\phi\sqrt{1+P_0^2}}{\sqrt{1+P_0^2}}$$

is a constant that depends on the initial value of the electron and positron momenta, P_0. We note that the coupled equations (9)–(13) depend on the sign of the linear momentum of the plasma; i.e., the plasma can move either in the positive or in the negative χ direction. It should be stressed that these equations describe the excitation of nonlinear relativistic upper-hybrid waves by neutrino beams in a magnetized plasma.

The nonrelativistic linear dynamics occurs for $P_{e,\bar e} \ll 1$, $\Psi \ll 1$, and $S_\nu \ll 1$. Subsequently, Eqs. (9)–(13) yield

$$\frac{d^2\Psi}{d\chi^2} = P_{ex} - P_{\bar e x} - S_\nu, \quad (14)$$

$$\frac{dP_{ey}}{d\chi} = -\Omega_c P_{ex}, \quad (15)$$

$$\frac{dP_{\bar e y}}{d\chi} = \Omega_c P_{\bar e x}, \quad (16)$$

$$\frac{d^2P_{ex}}{d\chi^2} + \Omega_{uh}^2 P_{\bar e x} = P_{\bar e x} + S_\nu, \quad (17)$$

$$\frac{d^2P_{\bar e x}}{d\chi^2} + \Omega_{uh}^2 P_{\bar e x} = P_{ex} - S_\nu, \quad (18)$$

where

$$\Omega_{uh}^2 = \omega^2/\omega_p^2 = 1 + \Omega_c^2$$

is the normalized squared upper-hybrid wave frequency. Equations (13)–(16) have been used to derive the coupled equations (17) and (18) for the x components of the pair-plasma momentum. These linear equations show that the pair-plasma dynamics is directly forced by the presence of the neutrino fluxes, which can generate strong electrostatic waves with an oscillating frequency ω, propagating perpendicularly to the external magnetic field direction.

Solving Eqs. (17) and (18) and substituting the results in Eq. (14), we obtain the following normalized

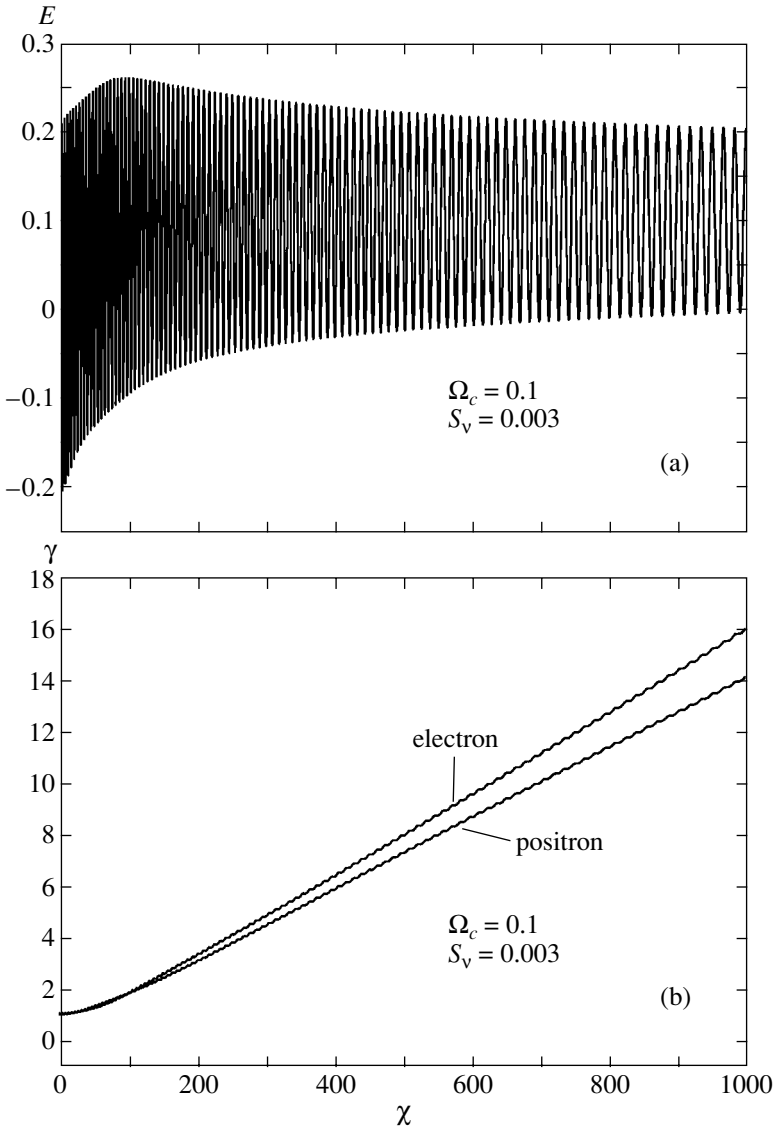

Fig. 1. Normalized electric field E (a) and normalized energy $\gamma = \sqrt{1 + P_x^2 + P_y^2}$ for pair plasmas (b) versus the normalized distance (phase) χ for the normalized gyrofrequency $\Omega_c = 0.1$ and the neutrino-driven term $S_\nu = 0.003$.

electric field ($E = -d\Psi/d\chi$) associated with the upper-hybrid wakefield:

$$E(x, t) = \frac{2\omega_p^3}{\omega^3} S_\nu \left(\frac{\Omega_c^2}{2}(kx - \omega t) + \sin(kx - \omega t) \right). \quad (19)$$

Here, the driven term S_ν is assumed constant according to our initial assumption, and the initial conditions $P_{e\bar{e}} = 0$, $\Psi = 0$, and $E = 0$ for $\chi = 0$, are imposed. As we can see, this field has a large amplitude, clearly dependent on the amount of neutrino energy deposited into the plasma, and is independent of the neutrino density; i.e., the plasma dynamics is triggered by the neutrino energy depositing into the plasma cloud. This strong electric field can accelerate plasma particles in the transverse direction, similarly to the surfatron acceleration mechanism [28], leading to the formation of a transverse pair-plasma jet.

4. NUMERICAL RESULTS

Figures 1a and 1b show the numerical solutions of Eqs. (9)–(13) for the neutrino-driven term $S_\nu = 3 \times 10^{-3}$, $\Omega_c = 0.1$, and $\gamma_\phi = 350$. In Fig. 1a, we display the normalized electric field associated with the excited upper-hybrid waves. This field reaches a maximum whose value is close to the normalized gyrofrequency, Ω_c. This occurs because the plasma particle momentum in the y direction is much larger than the momentum in the x direction due to the weakness of the neutrino term S_ν. We can also see that the nonlinear regime only starts for large phase numbers ($\chi > 750$). During the linear

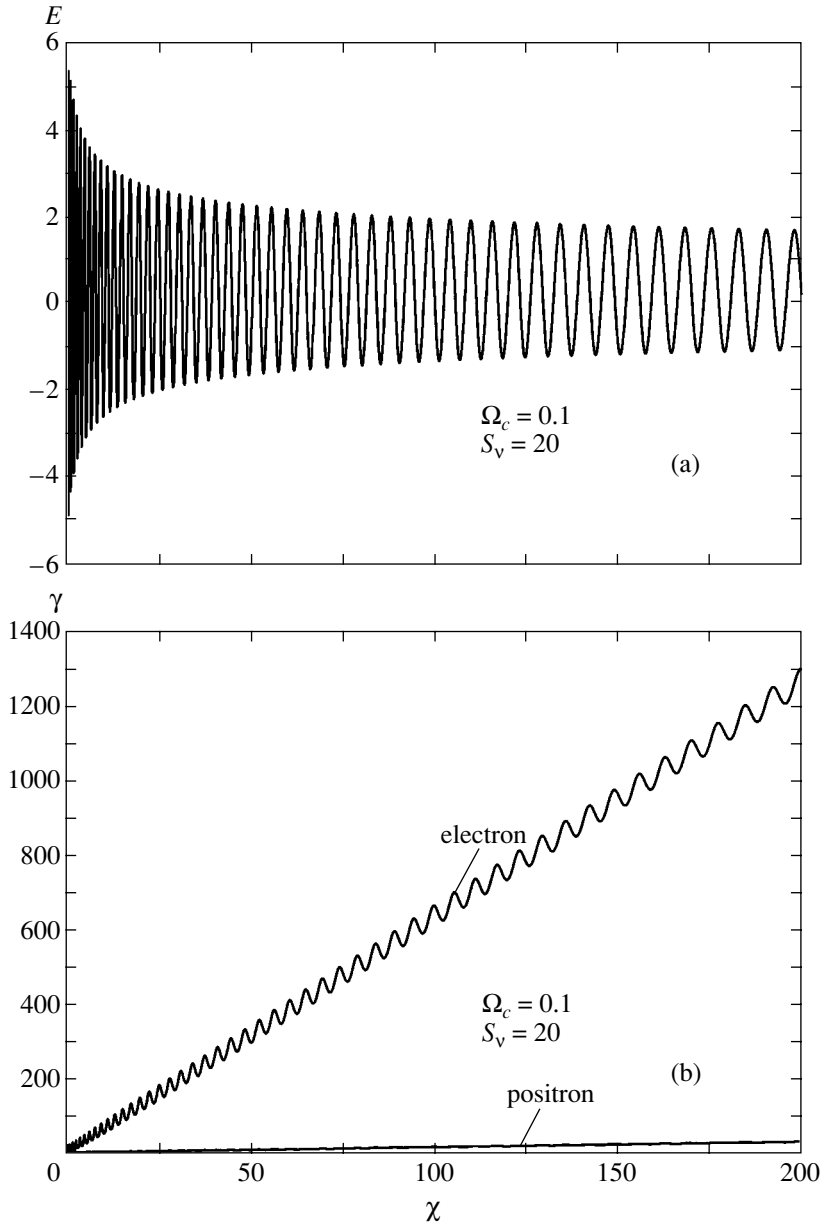

Fig. 2. Normalized electric field E (a) and normalized energy $\gamma = \sqrt{1 + P_x^2 + P_y^2}$ for a pair plasma (b) versus the normalized distance (phase) χ for the normalized gyrofrequency $\Omega_c = 0.1$ and the neutrino-driven term $S_\nu = 20$.

regime, the dynamics of the electron and positron particles of the plasma is symmetric, as expected. We note that this linear regime corresponds to the analytic solution given by Eq. (19). However, for long neutrino–plasma interactions, this symmetry is broken owing to the presence of the neutrino flux, see Fig. 1b.

Figure 2a shows the normalized wake electric field for $\Omega_c = 0.1$, as in Fig. 1a but with a more intense neutrino-driven term, $S_\nu = 20$. We observe that as the neutrino-driven term increases, the x component of the particle momentum becomes much larger than the y component, which brings the electric field to a saturation level much larger than the value shown in Fig. 1a. We also note that in this case, the electric field presents a high intensity at the beginning of the neutrino–plasma interaction, and after a while starts to decay and to transfer its energy to the plasma particles. After that, the field reaches a saturation value $E \approx 0.5$, which is maintained during the interaction. Therefore, the plasma particles feel a constant electric force, which leads to an unlimited transverse acceleration, as we can see from Fig. 2b. It should also be pointed out that in this case, the asymmetry between the electron and positron dynamics is very intense, due to the high value of the induced negative charge that the neutrinos acquire during the interaction with the pair plasma. This process

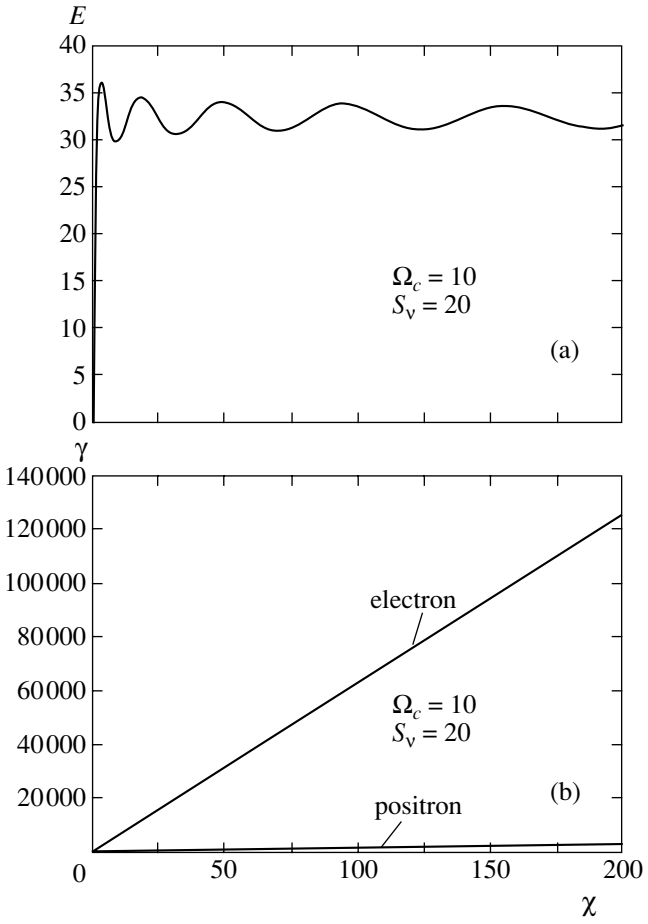

Fig. 3. Normalized electric field E (a) and normalized energy $\gamma = \sqrt{1 + P_x^2 + P_y^2}$ for a pair plasma (b) versus the normalized distance (phase) χ for the normalized gyrofrequency $\Omega_c = 10$ and the neutrino-driven term $S_\nu = 20$.

leads to a charge separation, which in turn creates the finite wakefields.

Until now, we have assumed that the gyrofrequency ω_c is smaller than the plasma frequency ω_p, which can occur in many astrophysical scenarios. Because our basic equations are normalized, we can also assume some astrophysical scenarios in which $\omega_c \gg \omega_p$, for example, a neutron star whose pair-plasma cloud surrounding the core of the star has a mean plasma number density close to $N_0 = 10^{30}$ cm^{-3} at 300 km away from the center of the star. This plasma density corresponds to the plasma frequency $\omega_p \approx 5.64 \times 10^{19}$ s^{-1}, which can be smaller than the gyrofrequency in some regions of the magnetized star.

Considering this scenario, we can assume the normalized gyrofrequency $\Omega_c = 10$, for instance. Figure 3a gives the normalized electric field associated with the upper-hybrid wakefield for the neutrino-driven term $S_\nu = 20$. We can see in this case that the electric field quickly reaches a huge saturation value, $E \approx 35$ (in terms of physical quantities, $E \approx 3.4 \times 10^{22}$ V/cm; we note that this value of the electric field is larger than the Schwinger limit for pair creation in vacuum) for a small value of the phase χ showing how intense the nonlinear behavior of the neutrino–plasma interaction is. In this extreme situation, plasma particles are unlimitedly accelerated to very high energy, as we can see from Fig. 3b. Of course, the symmetry between the electron and positron dynamics is completely destroyed (see, e.g., Fig. 3b) in the presence of these intense neutrino bursts, which means that the electron–plasmon cross section is now much larger than the positron–plasmon one.

5. CONCLUSIONS

In conclusion, we have presented a hydrodynamic description of large-amplitude upper-hybrid waves excited by intense neutrino beams in a dense magnetized pair plasma. The excitation of wakefields in a pair plasma is possible because of the spontaneous breakdown in symmetry between the electron and positron dynamics due to the driving force of intense neutrino bursts. Physically, the symmetry breaking is attributed to the induced negative charge that neutrinos acquire as they travel through the plasma. The induced negative charge pushes the electrons in such a way that the electron momentum increases and the positrons are attracted by the effectively charged neutrinos. The resulting charge imbalance due to the charge separation, in turn, creates finite-amplitude wakefields. Our results, which are independent of the neutrino density but dependent on the neutrino energy deposited into the plasma, are valid for any astrophysical scenarios. They should be applied to understand the acceleration of pairs in stars as well as in ultrarelativistic pair-plasma jets that produce γ-ray bursts. Furthermore, the presence of an external magnetic field shows that we have the possibility of generating upper-hybrid wakefields that can transversely accelerate the pair plasma to ultrarelativistic energies by the surfatron mechanism [28].

By means of numerical computation, we have observed that the effects of the finite plasma and neutrino temperatures do not significantly affect the pair-plasma dynamics and the amplitude intensity of the electric field generated by the neutrino bursts. It should be pointed out that any amount of neutrino energy transferred to the plasma suffices to drive the wakefields. Of course, the amplitude of the wakefield depends on the amount of neutrino energy deposited into the plasma. Hence, the intensity of the generated electric field can reach values that allow generating-pair creation by the wakefields. We are presently investigating pair creation in magnetized plasmas.

This research was partially supported by the CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) as CAPES (Coordenadoria de Apoio ao Pessoal do Ensino Superior), as well as by the European Commission (Brussels, Belgium) under contract no. HPRN-CT-2001-00314 and Sonderforschungsbe-

reich 591 of the Deutsche Forschungsgemeinschaft (Bonn, Germany). We also acknowledge the support provided by Centre for Fundamental Physics at the Rutherford Appleton Laboratory at Chilton, Didcot, UK.

REFERENCES

1. G. W. Gibbons, S. W. Hawking, and S. Siklos, *The Very Early Universe* (Cambridge Univ. Press, Cambridge, 1983).
2. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), p. 763.
3. F. C. Michel, Rev. Mod. Phys. **54**, 1 (1982).
4. T. Tajima and T. Taniuti, Phys. Rev. A **42**, 3587 (1990).
5. N. J. Shaviv and A. Dar, Astrophys. J. **447**, 863 (1995); A. Dar, Astrophys. J. **500**, L93 (1998).
6. H. R. Miller and P. J. Witta, *Active Galactic Nuclei* (Springer, Berlin, 1987).
7. F. T. Gratton, G. Gnavi, R. M. O. Galvâ, and L. Gamberoff, Phys. Rev. E **55**, 3381 (1997).
8. J. G. Lomonadze, L. Stenflo, V. N. Tsytovich, and H. Wilhelmsson, Phys. Scr. **26**, 455 (1982); A. C. L. Chian and C. F. Kennel, Astrophys. Space Sci. **97**, 9 (1983).
9. G. J. Fishman and C. A. Meegan, Annu. Rev. Astron. Astrophys. **33**, 415 (1995).
10. S. S. Gershtein, private communication; IHEP 2000-15.
11. P. K. Shukla, N. N. Rao, M. Y. Yu, and N. L. Tsintsadze, Phys. Rep. **138**, 1 (1986).
12. J. R. Wilson, in *Numerical Astrophysics*, Ed. by J. M. Centrella, J. M. LeBlanc, and R. L. Bowers (Bartlett and Jones, Boston, 1985), p. 422; H. A. Bethe and J. R. Wilson, Astrophys. J. **295**, 14 (1985).
13. J. Cooperstein, Phys. Rep. **163**, 95 (1988).
14. H. A. Bethe, Rev. Mod. Phys. **62**, 801 (1990).
15. S. Bludman, Da Hsuan Feng, Th. Gaisser, and S. Pittel, Phys. Rep. **256**, 3 (1995).
16. V. N. Oraevsky and V. B. Semikoz, Physica A (Amsterdam) **142**, 135 (1987).
17. J. F. Nieves and P. B. Pal, Phys. Rev. D **49**, 1398 (1994).
18. N. L. Tsintsadze, J. T. Mendonça, and P. K. Shukla, Phys. Lett. A **249**, 110 (1998).
19. J. T. Mendonça, A. Serbeto, P. K. Shukla, and L. O. Silva, Phys. Lett. B **548**, 63 (2002).
20. L. O. Silva, R. Bingham, J. M. Dawson, *et al.*, Phys. Rev. Lett. **83**, 2703 (1999).
21. A. Serbeto, Phys. Plasmas **6**, 2943 (1999).
22. A. J. Brizard and J. S. Wurtele, Phys. Plasmas **6**, 1323 (1999).
23. P. Meszaros, *High Energy Radiation from Magnetized Neutron Stars* (Univ. of Chicago Press, Chicago, 1992), Chap. 2.
24. G. Raffelt, *Stars as Laboratories for Fundamental Physics* (Univ. of Chicago Press, Chicago, 1996).
25. A. A. Gvozdev and I. S. Ognev, JETP Lett. **74**, 298 (2001).
26. J. Pantalone, Phys. Lett. B **342**, 250 (1995).
27. A. Serbeto, L. A. Rios, and P. K. Shukla, Phys. Plasmas **9**, 4406 (2002).
28. T. Katsouleas and J. M. Dawson, Phys. Rev. Lett. **51**, 392 (1983).

Journal of Experimental and Theoretical Physics, Vol. 99, No. 3, 2004, pp. 474–479.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 3, 2004, pp. 542–548.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Renormalons and Analytic Properties of the β Function

I. M. Suslov

Kapitza Institute of Physical Problems, Russian Academy of Sciences, Moscow, 117334 Russia
e-mail: suslov@kapitza.ras.ru

Received April 19, 2004

Abstract—The presence or absence of renormalon singularities in the Borel plane is shown to be determined by the analytic properties of the Gell-Mann–Low function $\beta(g)$ and some other functions. A constructive criterion for the absence of singularities consists in the proper behavior of the β function and its Borel image at infinity, $\beta(g) \propto g^\alpha$ and $B(z) \propto z^\alpha$ with $\alpha \le 1$. This criterion is probably fulfilled for the φ^4 theory, quantum electrodynamics, and quantum chromodynamics, but is violated in the $O(n)$-symmetric sigma model with $n \longrightarrow \infty$.

(1) More than twenty years ago, Lipatov [1] suggested a method for calculating the high orders of perturbation theory according to which these are determined by the saddle-point configurations (instantons) of the corresponding functional integrals. The method proved to be applicable to a wide range of problems [2, 3], but it was soon questioned in connection with the detection of factorially large contributions from the individual diagrams, renormalons [4]. In the opinion of 't Hooft [5], the latter are not contained in the instanton contribution. Formally, the asymptotics of perturbation theory is determined by the singularity in the Borel plane closest to the coordinate origin. Whereas the presence of instanton singularities is beyond doubt, the existence of renormalon singularities has never been proven, which is recognized by the most active proponents of this trend [6]: such singularities can be easily obtained by the summation of individual sequences of diagrams, but it cannot be made sure that these are preserved when all diagrams are taken into account. Previously [7], we proposed a proof of the absence of renormalon singularities in the φ^4 theory, which calls into question the idea of renormalons as a whole; however, there is no similar proof for other field theories. The analysis performed below clarifies the situation with renormalon singularities in an arbitrary field theory: in general, their presence or absence is determined by the analytic properties of the Gell-Mann–Low function and some other functions.

The simplest class of renormalon diagrams arises in quantum electrodynamics after the separation of an internal photon line in an arbitrary diagram (Fig. 1a) and the insertion of a chain of electron loops into it (Fig. 1b). In the original diagram, the integration $\int d^4 k k^{-2n}$ over the range of large momenta ($n = 3, 4, \ldots$) corresponded to the separated photon line with momentum k. When N electron loops are inserted into the photon line, the additional factor $\ln^N(k^2/m^2)$ (m is the electron mass) emerges in the integrand; the integration yields a quantity on the order of $N!$. Arbitrary insertions into the photon line lead to the substitution of the running coupling constant $g(k^2)$ for the interaction constant g_0 (Fig. 1c) and give rise to the integral $\int d^4 k k^{-2n} g(k^2)$. The summation of the chains of loops corresponds to using the single-loop approximation $\beta(g) = \beta_2 g^2$ for the Gell-Mann–Low function and yields the well-known result

$$g(k^2) = \frac{g_0}{1 - \beta_2 g_0 \ln(k^2/m^2)}. \qquad (1)$$

After the integration over $k^2 \gtrsim m^2$, we obtain

$$\int d^4 k k^{-2n} g(k^2) = g_0 \sum_N \int d^4 k k^{-2n} \left(\beta_2 g_0 \ln \frac{k^2}{m^2}\right)^N \sim g_0 \sum_N N! \left(\frac{\beta_2}{n-2}\right)^N g_0^N. \qquad (2)$$

After the Borel summation, this yields renormalon sin-

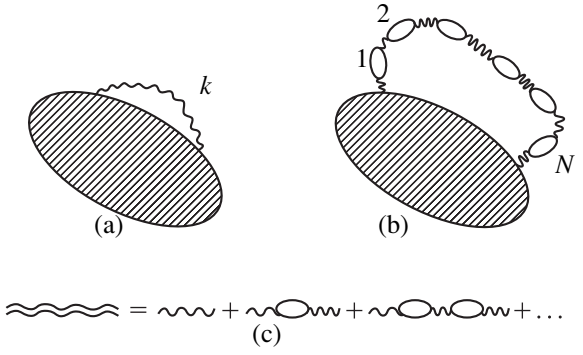

Fig. 1.

gularities at the points[1]

$$z_n = \frac{n-2}{\beta_2}, \quad n = 3, 4, 5, \ldots \tag{3}$$

in the Borel z plane.

The analysis performed below is based on the fact that for a given β function, the summation of the entire class of diagrams obtained by all the possible insertions into the photon line presents no problem: it will suffice to solve the Gell-Mann–Low equation

$$\frac{dg}{d\ln k^2} = \beta(g) = \beta_2 g^2 + \beta_3 g^3 + \ldots \tag{4}$$

with the initial condition $g(k^2) = g_0$ at $k^2 = m^2$ and to analyze the expansion in terms of g_0 for an integral of type (2). The more complex classes of renormalon diagrams can be studied by using the general renormalization group equation in Callan–Symanzik form.

(2) As an illustration, let us consider the model β function

$$\beta(g) = \frac{\beta_2 g^2}{1+g^2}, \tag{5}$$

for which Eq. (4) can be easily solved:

$$g(k^2) = -\frac{1}{2}\left(\frac{1}{g_0} - g_0 - x\right) + \sqrt{\frac{1}{4}\left(\frac{1}{g_0} - g_0 - x\right)^2 + 1}, \tag{6}$$

where $x = \beta_2 \ln(k^2/m^2)$. The right-hand side is regular at $g_0 = 0$ and can be expanded into a power series of g_0. The structure of this series is

$$g = \sum_{N=1}^{\infty} A_N \left\{\frac{g_0}{r(x)}\right\}^N, \tag{7}$$

where $r(x)$ is the radius of convergence and the coefficients A_N depend on N as a power law. The radius of convergence is determined by the distance to the singularity closest to the coordinate origin. The singularities g_c on the right-hand side of Eq. (6) correspond to zeros of the radicand and are defined by the equation

$$g_c^2 + (x + 2i)g_c - 1 = 0 \tag{8}$$

and its complex conjugate equation. At large x, the minimum (in absolute value) root is $g_c \approx 1/x$ and series (7) takes the form

$$g(k^2) = \sum_{N=1}^{\infty} A_N (g_0 x)^N = \sum_{N=1}^{\infty} A_N \left(\beta_2 \ln\frac{k^2}{m^2}\right)^N g_0^N. \tag{9}$$

Substituting it into integral (2) yields singularities at points (3) (at large N, the integral is determined by large k that correspond to large x). Thus, Parisi's reasoning [8] that renormalon singularities can exist in internally consistent theories is confirmed.

(3) The overall picture is determined by the fate of the Landau pole in the single-loop result (1). This pole can remain on the real axis, shift into the complex plane, or go to infinity. The right-hand side of Eq. (1) as a function of g_0 changes on the characteristic scale $(\ln(k^2/m^2))^{-1}$; this property does not change when the higher loops are taken into account, because result (1) is always valid at small g_0. If $g(k^2)$ as a function of g_0 has singularities in a finite part of the complex plane, then the characteristic scale of its change is naturally determined by the distance to the closest singularity, which thus proves to be on the order of $(\ln(k^2/m^2))^{-1}$, generating a series of type (9) and renormalon singularities. However, this is not always the case: for example, the characteristic scale of the change for integer functions is determined by other factors, and the above conclusion ceases to be valid.

The general solution of the Gell-Mann–Low equation is

$$F(g) = F(g_0) + \ln\frac{k^2}{m^2}, \tag{10}$$

where

$$F(g) = \int \frac{dg}{\beta(g)}.$$

Taking into account the behavior of the function $F(g)$ at small g, we can write

$$F(g) = -\frac{1}{\beta_2 g} + f(g), \quad \text{where } \lim_{g\to 0} g f(g) = 0, \tag{11}$$

and, formally resolving (10) for g, obtain

$$g(k^2) = F^{-1}\left\{-\frac{1}{\beta_2 g_0} + f(g_0) + \ln\frac{k^2}{m^2}\right\}. \tag{12}$$

If the function $z = F(g)$ is regular at g_0 and $F'(g_0) \neq 0$, then the inverse function $g = F^{-1}(z)$ that is also regular exists in some neighborhood of the point g_0. Therefore, the singularities of the function $F^{-1}(z)$ are $z_c = F(g_c)$, where all the possible g_c are defined by the condition

$$F'(g_c) = 0 \text{ or } F'(g_c) \text{ does not exist.} \tag{13}$$

The singularities in variable g_0 in (12) are defined by the equation

$$z_c = -\frac{1}{\beta_2 g_0} + f(g_0) + \ln\frac{k^2}{m^2} \tag{14}$$

or

$$g_0 x - 1 = \beta_2 g_0 [z_c - f(g_0)]. \tag{15}$$

If z_c is finite, then Eq. (15) at large x has a root $g_0 \approx 1/x$ in the range of small g_0, where the right-hand side of

[1] Similar singularities with $n = 0, -1, -2, \ldots$ arise from the integration over the range of small momenta (infrared renormalons).

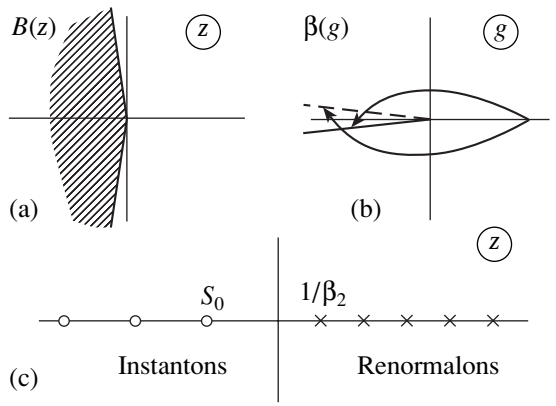

Fig. 2. It follows from the analyticity of the function $B(z)$ at $|\arg z| < \pi/2 + \delta$ (a) that $\beta(g)$ is analytic at $|\arg z| < \pi + \delta$ (b), i.e., on the entire physical sheet of the Riemann surface; (c) the picture of singularities in the Borel plane for the φ^4 theory and quantum electrodynamics suggested by 't Hooft (S_0 is the minimum instanton action, β_2 is the first nonvanishing expansion coefficient of the β function).

Eq. (15) is insignificant in view of (11). Thus, there is a singularity at $g_c \approx 1/x$ that generates series (9) leading to the renormalon singularities (3). If, alternatively, $z_c = \infty$, then Eq. (14) has no solutions for $g_0 \sim 1/x$, and an expansion of type (9) is possible only with the coefficients A_N that decrease faster than any exponential: the renormalon contribution is definitely much smaller than the instanton contribution, and no singularities emerge in the Borel plane. The solutions with $g_0 \sim 1$ (which are possible due to the singularities of the function $f(g_0)$) are definitely unrelated to the renormalon mechanism: their contribution is determined by a series in which g_0^N is not accompanied by a factor of the type $(\ln(k^2/m^2))^N$. To summarize, we have reached the following conclusion. Renormalon singularities take place if there exists at least one point g_c (including $g_c = \infty$) for which condition (13) is satisfied and $z_c = F(g_c) < \infty$; otherwise, no renormalon singularities exist.

It remains to reformulate the results in terms of the β function itself. First, note that a regular root of the form

$$\beta(g) \propto (g - g_c)^n, \quad n = 1, 2, 3, \ldots$$

does not lead to renormalons: in this case, the derivative $F'(g_c)$ does not exist, but $F(g_c) = \infty$; in particular, this is true for the root at $g = 0$. A power-law behavior at infinity, $\beta(g) \propto g^\alpha$, generates renormalons only at $\alpha > 1$ (which corresponds to the existence condition for the Landau pole of a nonalternating function $\beta(g)$). All of the other possibilities for the satisfaction of condition (13) are related to the singularities of the function $\beta(g)$ at finite points g_c: for renormalons to exist, these must be strong enough for the function $1/\beta(g)$ to be integrable at g_c (e.g., $\beta(g) \propto (g - g_c)^\gamma$ with $\gamma < 1$). A sufficient condition for the absence of renormalons is the regularity of the function $\beta(g)$ at finite g and its power-law behavior, $\beta(g) \propto g^\alpha$ with $\alpha \le 1$, at infinity; in fact, weak singularities of the type $\beta(g) \propto (g - g_c)^\gamma$ with $\gamma > 1$ are admissible at finite g.

(4) If all of the singularities in the Borel plane are assumed to be of instanton or renormalon origin,[2] then a constructive criterion for the absence of renormalon singularities can be formulated.

The perturbative series for the β function is fractorially divergent [1–3], because there is a cut in the complex g plane that emerges from the coordinate origin. Therefore, $g = 0$ is the branching point, as is generally $g = \infty$. The function $\beta(g)$ is represented by the Borel integral

$$\beta(g) = \int_0^\infty dz e^{-z} B(gz) = g^{-1} \int_0^\infty dz e^{-z/g} B(z), \qquad (16)$$

where $B(z)$ is the Borel image of the function $\beta(g)$. Let us assume that it has a power-law behavior at infinity, $B(z) \propto z^\alpha$ (then $\beta(g) \propto g^\alpha$), and is regular for $|\arg z| < \pi/2 + \delta$, $\delta > 0$ (Fig. 2a). Directing the contour of integration along the ray $z = |z| e^{i\phi_0}$, we can easily verify that integral (16) converges for $g = |z| e^{i\phi_0}$ with $|\phi - \phi_0| < \pi/2$. Since the contour can turn through angles $|\phi_0| < \pi/2 + \delta$, the function $\beta(g)$ is regular for $|\arg g| < \pi + \delta$ (Fig. 2b), implying that there are no singularities at finite points on the physical sheet of the Riemann surface. In this case, the behavior of the β function at infinity ($\beta(g) \propto g^\alpha$ with $\alpha \le 1$) yields the condition for the absence of renormalon singularities.

The derived criterion can be constructively used as follows. Consider the φ^4 theory or quantum electrodynamics; in this case, there are instanton singularities on the negative semiaxis and, possibly [5], renormalon singularities on the positive semiaxis (Fig. 2c). Let us assume that there are no renormalon singularities. In this case, (i) the regularity condition for the function $\beta(g)$ at finite g (Figs. 2a and 2b) is satisfied; (ii) the asymptotic form of the expansion coefficient β_N is determined by the nearest instanton singularity and can be found by Lipatov's method; (iii) the Borel integral is well defined, and the perturbative series for the function $\beta(g)$ admits an unambiguous summation, which allows its behavior at infinity to be determined. If the β function increases faster than g^α with $\alpha > 1$, then the initial assumption is invalid, and the existence of renormalon

[2] This assumption has not been rigorously proven, but nobody has proposed a viable alternative to it. It can be justified by the fact that all of the singularities in the Borel plane for finite-dimensional integrals are related to the extrema of the action (in this case, the reasoning of 't Hooft in [5] is necessary and sufficient), while the renormalon singularities are explicitly related to the passage to an infinite number of integrations.

singularities has been proven by contradiction. If, alternatively, $\beta(g) \propto g^\alpha$ with $\alpha \leq 1$, then the assumption about the absence of renormalon singularities is self-consistent.

The outlined program for the above theories was carried out previously [9, 10] by interpolating Lipatov's asymptotics with known values of the first expansion coefficients and yielded $\alpha = 0.96 \pm 0.01$ for the φ^4 theory [9] and $\alpha = 1.0 \pm 0.1$ for quantum electrodynamics [10]. Thus (within the uncertainty of the results), the self-consistent exclusion of renormalon singularities proves to be possible. Moreover, a comparison with existing analytic estimates is indicative of the exact equality $\alpha = 1$ in both cases [9, 10]. In any case, the β functions in these theories are nonalternating ones[3] and the condition for the absence of renormalon singularities in them is identical to the condition for their internal consistency.

For quantum chromodynamics, $\alpha = -12 \pm 3$ [11] and the instanton singularities lie on the positive semiaxis. The assumption about the absence of renormalon singularities is self-consistent for the sheet of the Riemann surface obtained by the analytic continuation from negative g (the sign of the Borel image changes as the sign of g changes, and the singularities pass to the negative semiaxis); this is enough to justify the procedure for determining[4] the index α used in [11]. The Borel integral at positive g must be properly interpreted to establish a connection with the physical sheet (its principal-value interpretation is not always correct [12]).

The only field theory in which the existence of renormalon singularities is deemed to have been firmly established is the $O(n)$-symmetric sigma model in the limit $n \longrightarrow \infty$ [6]. In this case, the single-loop β function is exact and $\beta(g) \propto g^2$ for $g \longrightarrow \infty$; consequently, $\alpha = 2$ and the self-consistent exclusion of renormalons proves to be impossible. However, this theory is internally inconsistent in the four-dimensional case.

Curiously, according to the formulated criterion, the truncation of the series for the β function at any finite number of terms immediately creates renormalon singularities. This shows that the problem of renormalons cannot be solved in terms of the loop expansion [13].

Note that the possibility of the existence of renormalon singularities makes the functional integrals ill-defined. The classical definition of the functional integral via the perturbation theory is defective, because the expansion in terms of the coupling constant is divergent: its constructive summation requires knowing the analytic properties in the Borel plane that are uncertain until it is established whether the renormalon singularities exist. It is also doubtful that the definition of the functional integral as a multidimensional integral on a lattice is correct: the lattice theory can differ fundamentally from the continuum theory, because the renormalon contributions are determined by the range of arbitrarily large momenta. An impasse is reached: the solution of the problem of renormalons requires studying the functional integrals, while the latter are ill-defined because the problem of renormalons is unsolved. The proposed scheme for the self-consistent exclusion of renormalon singularities is probably the only possible way out of the situation. In this case, the continuum theory, by definition, is understood to be the limit of the lattice theories.

(5) In general, a certain class of renormalon diagrams is singled out by the condition that new vertices are inserted into the same element (a line or a vertex) of the original skeleton diagram. This definition allows the existence conditions for the main renormalon contribution to be analyzed: if new vertices are inserted with an equal probability into m different elements, then the corresponding contribution is on the order of $[(N/m)!]^m \sim N!m^{-N}$ and contains the redundant smallness m^{-N}.[5] For electrodynamics, integral (2) considered above corresponds to the summation of the class of diagrams obtained by all the possible insertions into the same photon line. A similar integral for the φ^4 theory corresponds to all the possible loop insertions to the same vertex; the domain of integration in which all momenta of the derived four-leg vertex are of the same order of magnitude is considered. In general, the renormalon integral is

$$\int d^4k k^{-2n} \Gamma(g_0, k), \tag{17}$$

where $\Gamma(g_0, k)$ is the vertex with M external lines from which M' lines carry a large momentum, on the order of k. The dependence on k is defined by the Callan–Simanchik equation

$$\left[-\frac{\partial}{\partial \ln k^2} + \beta(g_0)\frac{\partial}{\partial g_0} + \gamma(g_0)\right]\Gamma(g_0, k) = 0, \tag{18}$$

where $\gamma(g_0)$ depends on M and M'. The general solution of Eq. (18) is

$$\Gamma(g_0, k) = \exp F_1(g_0)\Phi\left(F(g_0) + \ln\frac{k^2}{m^2}\right),$$
$$F_1(g) = -\int dg \frac{\gamma(g)}{\beta(g)}, \tag{19}$$

[3] For the φ^4 theory, we have in mind the four-dimensional case in which the problem of renormalons is of current interest.

[4] Note that the asymptotics $\beta(g) \propto g^\alpha$ does not guarantee a power-law behavior of the Borel image, $B(z) \propto z^\alpha$, in all directions in the complex plane (e.g., $B(z) = \beta_2(1 - \cos z)$ for the model β function (5)). Therefore, we emphasize that the index α in [9–11] was determined directly from the asymptotics of the Borel image, and the assumption about its power-law behavior was subjected to a special test.

[5] In fact, in this case, there are m independent integrations of type (17) for each of which the condition for the absence of renormalon contributions is identical to that established below.

where $\Phi(z)$ is an arbitrary function. If z_c is a singularity of the function $\Phi(z)$, then the singularities in variable g_0 are defined by Eq. (14). The function $\Phi(z)$ can be expressed in terms of $R(g_0) \equiv \Gamma(g_0, m)$,

$$\Phi(z) = R(F^{-1}(z))\exp\{-F_1(F^{-1}(z))\}, \qquad (20)$$

and the singularities of the function $F^{-1}(z)$ are those of the function $\Phi(z)$. Therefore, the condition for the existence of renormalons found above is also sufficient in the general case. Additional possibilities for their emergence are associated with the singularities of the functions $F_1(g)$ and $R(g)$. If one of them is singular at g_c, then $z_c = F(g_c)$ is a singularity of the function $\Phi(z)$. The functions $F_1(g)$ and $R(g)$ are represented by Borel integrals of type (16) and have $g = 0$ and ∞ as the branching points. However, this does not lead to singularities of the function $\Phi(z)$ at finite points, because

$$F(0) = \infty, \quad F(\infty) = \infty \ (\text{for } \alpha \le 1).$$

At the same time, the singularities at finite g can be self-consistently excluded for the functions $F_1(g)$ and $R(g)$, as was done above for the β function. As a result, the behavior of the function $\beta(g)$ at infinity also determines the presence or absence of renormalons in the general case.

(6) It is clear from the above discussion that using information only from the renormalization group, we can establish the necessary and sufficient conditions for the existence of renormalons, but cannot reach any definite conclusions. Let us compare this with Parisi's renormalization-group analysis [8] that underlies all of the recent studies devoted to renormalons [6]. If, following [8], the momentum dependence of the Borel images is assumed to differ from the single-loop result only by a slowly changing factor, then this ansatz formally satisfies the equations if we expand the slowly changing function in terms of gradients and restrict our analysis to the local approximation. However, to study the stability of the solution, we must continue the expansion in terms of gradients and obtain a diffusion-type equation. The solution is stable if the corresponding diffusion coefficient is positive, which, in general, is not the case. The extent to which Parisi's solution breaks down is determined by the rather subtle properties of the β function, which correlates with the assertions of this work.

In conclusion, let us discuss the subtle point in the proof of the φ^4 theory that was inadequately covered in [7]. Any quantity defined by the perturbative series is a function of the bare charge g_B and the cutoff parameter Λ. Passing to the renormalized charge g gives rise to the function $F(g, \Lambda)$ that contains the residual dependence on Λ, but has the following finite limit in view of the renormalizability:

$$\lim_{\Lambda\to\infty} F(g, \Lambda) = F(g). \qquad (21)$$

A similar property is expected for the corresponding Borel images:

$$\lim_{\Lambda\to\infty} B(z, \Lambda) = B(z). \qquad (22)$$

The analyticity of the function $B(z, \Lambda)$ at finite Λ in the complex z plane with a cut from the first instanton singularity to infinity was rigorously proven previously [7]. The function $B(z)$ is analytic in the same domain on the condition of uniform convergence in (22) (the Weierstrass theorem [14]); the latter takes place if the function $B(z, \Lambda)$ is bounded (the compactness principle for regular functions [15]). Therefore, the finiteness of the limit in (22) is enough to prove[6] the regularity of the function $B(z)$.

Unfortunately, the finiteness of the limits in (21) and (22) has been rigorously proven only in the framework of the perturbation theory, i.e., not for the functions $F(g, \Lambda)$ and $B(z, \Lambda)$ themselves, but for the coefficients of their expansion in powers of g and z. The proof in [7] suggests that the limits are finite at the level of the functions, and, in this sense, it is incomplete. However, the finiteness of the limits in (21) and (22) is required for the existence of renormalizability and must be considered as a necessary physical condition for it. The latter is closely related to the necessity of redefining the functional integrals noted above.

ACKNOWLEDGMENTS

I am grateful to L.N. Lipatov for numerous discussions on the problem of renormalons during which the idea of this work arose. This study was supported by the Russian Foundation for Basic Research (project no. 03-02-17519).

REFERENCES

1. L. N. Lipatov, Zh. Éksp. Teor. Fiz. **72**, 411 (1977) [Sov. Phys. JETP **45**, 216 (1977)].
2. *Large-Order Behavior of Perturbation Theory*, Ed. by J. C. Le Guillou and J. Zinn-Justin (North-Holland, Amsterdam, 1990).

[6] If the function $B(z)$ has a singularity at z_0 that is the point of regularity for $B(z, \Lambda)$, then the function $B(z, \Lambda)$ is unbounded for $\Lambda \longrightarrow \infty$ on any contour that encloses point z_0. This unboundedness is global in nature and is unrelated to the possible divergence of the function $B(z)$ at $z = z_0$. In fact, if the Borel image of $B(z)$ becomes infinite at isolated points in the complex plane, then it is easy to ensure its boundedness for the case of power-law divergence of relevant interest. Let, by contradiction, $B(z) \propto (z - z_0)^{-\gamma}$ in the neighborhood of z_0 that is the point of regularity for $B(z, \Lambda)$. Let us turn to a more general definition of the Borel image in which the coefficients of the initial series are divided not by $N!$, but by $\Gamma(N + b_0)$; the index γ can then be made negative by increasing b_0 [7, Section 3.1]. Thus, the function $B(z)$ becomes bounded near z_0, and no singularity can exist at $z = z_0$. The regularity at arbitrary b_0 follows from the regularity of $B(z)$ at large b_0 [7, Section 3.1].

3. E. B. Bogomolny, V. A. Fateyev, and L. N. Lipatov, in *Soviet Science Reviews, Section A: Physics Reviews*, Ed. by I. M. Khalatnikov (Harwood Academic, New York, 1980), Vol. 2, p. 247.
4. B. Lautrup, Phys. Lett. B **69B**, 109 (1977).
5. G. 't Hooft, in *The Whys of Subnuclear Physics, Erice, 1977*, Ed. by A. Zichichi (Plenum, New York, 1979).
6. M. Beneke, Phys. Rep. **317**, 1 (1999).
7. I. M. Suslov, Zh. Éksp. Teor. Fiz. **116**, 369 (1999) [JETP **89**, 197 (1999)].
8. G. Parisi, Phys. Rep. **49**, 215 (1979).
9. I. M. Suslov, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 315 (2000) [JETP Lett. **71**, 217 (2000)].
10. I. M. Suslov, Pis'ma Zh. Éksp. Teor. Fiz. **74**, 211 (2001) [JETP Lett. **74**, 191 (2001)].
11. I. M. Suslov, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 387 (2002) [JETP Lett. **76**, 327 (2002)].
12. R. Seznec and J. Zinn-Justin, J. Math. Phys. **20**, 1398 (1979).
13. F. David, Nucl. Phys. B **209**, 433 (1982); **234**, 237 (1984); **263**, 637 (1986).
14. G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers*, 2nd ed. (McGraw-Hill, New York, 1968; Nauka, Moscow, 1977).
15. M. A. Evgrafov, *Analytic Functions*, 2nd ed. (Nauka, Moscow, 1968; Saunders, Philadelphia, Pa., 1966).

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 99, No. 3, 2004, pp. 480–486.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 3, 2004, pp. 549–557.

ATOMS, SPECTRA, RADIATION

Enhancement of Resonant Bleaching of *J*-aggregates upon Lengthening of an Exciting Radiation

R. V. Markov[a,*], A. I. Plekhanov[a], Z. M. Ivanova[c], N. A. Orlova[b], V. V. Shelkovnikov[b], A. A. Ivanov[c], and M. V. Alfimov[c]

[a] *Institute of Automatics and Electrometry, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia*
[b] *Novosibirsk Institute of Organic Chemistry, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia*
[c] *Research Center of Photochemistry, Russian Academy of Sciences, Moscow, 117421 Russia*
**e-mail: fractal@iae.nsk.su*

Received February 26, 2004

Abstract—Thin films of *J*-aggregates of a new amphiphilic thiacarbocyanine dye of the benthiazole series are prepared and the nonlinear optical response of molecular *J*-aggregates is studied for femto- and nanosecond exciting radiation pulses. It is found that the nonlinear optical response of *J*-aggregates exhibits substantial enhancement upon an increase in the pulse duration, which cannot be described by the saturation effect in the model of a two-level system. This effect is considered using a three-level model taking into account the formation of self-trapped exciton states in molecular *J*-aggregates. *© 2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Ultrafast nonlinear optical transformation of signals is based on third-order nonlinear optical effects, for which the absorption coefficient and the refractive index of the medium are functions of the radiation intensity. The effective transformation of signals requires media with high and rapidly varying nonlinear-optical coefficients (with a response time shorter than 1 ps) as well as high thermal and photochemical stability. In this connection, nanostructural materials (such as semiconductor quantum wells and dots, polyconjugate polymers, fullerenes, and metallic clusters) in which the combination of required properties is ensured by their dimension and discreteness, have become the objects of intense studies in recent years.

Among promising materials, the class of molecular aggregates or *J*-aggregates of organic dye molecules deserves special attention [1]. A distinguishing feature of *J*-aggregates is that aggregation leads to the emergence of a narrow absorption band displaced towards low frequencies relative to the absorption band of nonaggregated molecules (monomers). The optical properties of *J*-aggregates are successfully described by the Frenkel model of excitons (see, for example, [2, 3]). Recent theoretical and experimental studies revealed that molecular *J*-aggregates exhibit a high third-order resonant nonlinear-optical susceptibility [4–7].

J-aggregates are self-organized quasi-one-dimensional ordered structures consisting of dye molecules. Translational symmetry and the interaction between molecules delocalize the excitation of an individual molecule over several coherently coupled molecules. Delocalization of excitation in turn leads to a unique combination of optical properties of *J*-aggregates: an intense and narrow absorption line [2, 3]; effective transmission of the absorbed energy to various acceptors (see [1], pp. 1–40 and 199–208); giant optical nonlinearity [8, 9]; and ultrashort relaxation time for the excited state [10]. The existence of these properties opens broad prospects for a number of practical applications of molecular *J*-aggregates, including the terrahertz demultiplexing of optical signals [11], passive mode locking for obtaining ultrashort laser pulses [12, 13], spectral sensibilization in photography (see [1], pp. 209–228), energy transfer from a light-harvesting antenna in systems of artificial photosynthesis [14, 15], the development of logic elements on the basis of mirrorless optical bistability [16]; and the formation of narrow-band luminescent organic composites with electroconducting polymers [17].

The nonlinear optical response of molecular *J*-aggregates of various organic dyes was investigated with the help of femto- [1, 18, 19], pico- [20, 21], and nanosecond [5, 22, 23] laser pulses. However, the investigations devoted to comparison of the nonlinear optical response of molecular *J*-aggregates under the action of laser pulses of various duration have not been carried out as yet.

This study is devoted to analysis of the optical nonlinearity of molecular *J*-aggregates of a new amphiphilic thiacarbocyanine dye under the action of femtosecond as well as nanosecond laser pulses.

2. SAMPLE PREPARATION AND EXPERIMENTAL CONDITIONS

To obtain *J*-aggregates in thin films, we used amphiphilic thiacarbocyanine dye with the structural

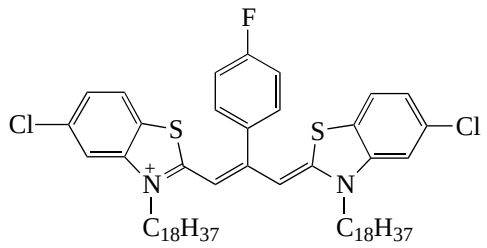

Fig. 1. Structural formula of thiacarbocyanine dye.

formula shown in Fig. 1. The synthesis of this dye is described in [24].

Films of *J*-aggregates were prepared on glass substrates by spin-coating of a solution of thiacarbocyanine dye or a dye–polymer (polymethyl metacrylate) mixture in the weight ratio 1 : 10 and concentration $c = 5 \times 10^{-3}$ mole/l at a speed of rotation from 1000 to 3000 rpm. A mixture of acetonitryl, dichlorethane, and chloroform in the volume ratio 2 : 2 : 1 was used as a solvent. The thickness of the dye film deposited on the substrate was about 30 nm. The thickness of the polymer film containing *J*-aggregates was 1100 nm. The thickness and optical constants n_0 and k of the complex refractive index of thin films were measured with the help of the LEF-752 ellipsometer manufactured at the Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences. The error of ellipsometric measurements of the film thickness was not larger than 0.5 nm. The thickness of films with a polymer matrix was measured using an interference microscope MII-4 to within 10 nm.

Figure 2 shows the characteristic absorption spectra of the films obtained from a dye solution and a solution of the dye–polymer mixture. The sharp peak near 630 nm corresponds to exciton absorption of *J*-aggregates (*J* peak). The high-frequency wing (in Fig. 2a) and peaks (in Fig. 2b) in the region of 585 nm are due to absorption of dye monomers. Solid curves in Fig. 2 correspond to the spectra measured for freshly prepared samples, while dashed curves correspond to the spectra measured after eight months. It can be seen that the prepared films hardly change their properties in the region of the *J* peak over a long time.

The ellipsometric measurements of the complex refractive index $n = n_0 - ik$ were carried out at the wavelength $\lambda = 632.8$ nm of a He–Ne laser. For dye films deposited on a glass substrate, these values were $n_0 = 2.69 \pm 0.02$ and $k = 1.8 \pm 0.1$, which corresponds to the absorption coefficient $\alpha_0 = 4\pi k/\lambda = 3.6 \times 10^5$ cm^{-1}.

The main characteristic of the resonant nonlinear optical response studied in our experiments is the nonlinear absorption coefficient β. The value of this coefficient is proportional to the imaginary part of the third-order nonlinear optical susceptibility and can be deter-

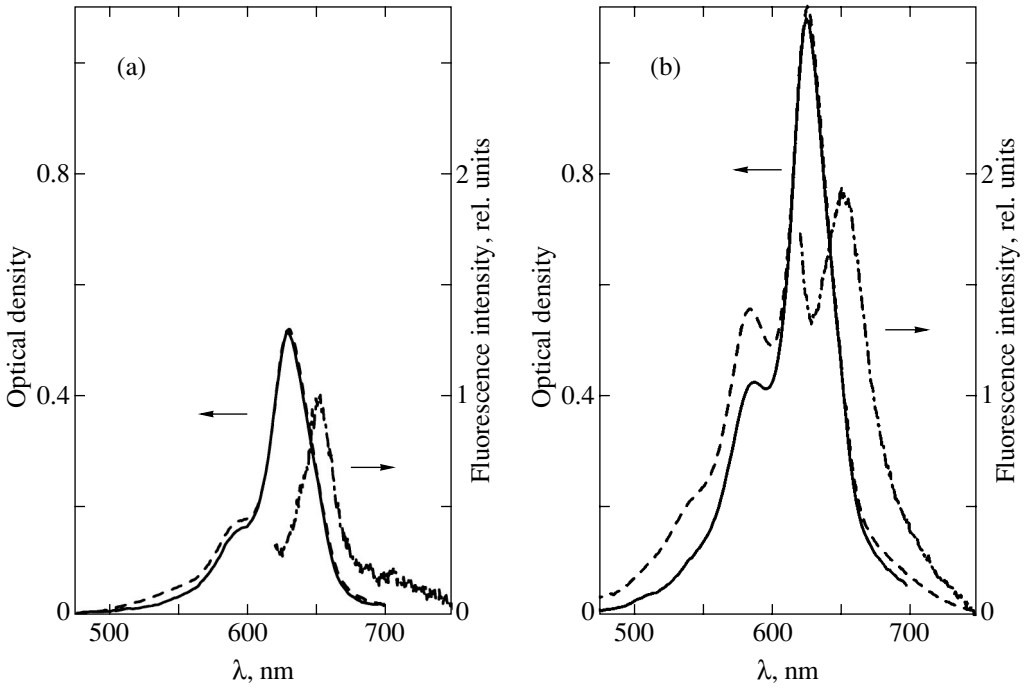

Fig. 2. Absorption spectra of films of *J*-aggregates of thiacarbocyanine: (a) dye film of thickness $L = 30$ nm, deposited on a glass substrate; (b) polymer film containing a dye, $L = 1100$ nm, the dye : polymer weight ratio is 1 : 10. The dashed curves correspond to the spectra of freshly prepared samples and the solid curves are the spectra of the same samples recorded after 8 months. The dot-and-dash curves correspond to fluorescence spectra of polymer films containing *J*-aggregates of thiacarbocyanine at room temperature. The wavelength of exciting radiation is 580 nm.

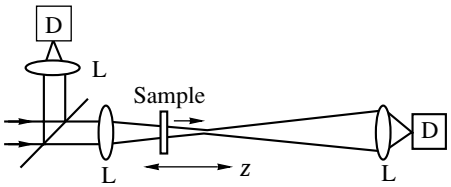

Fig. 3. Schematic diagram illustrating the longitudinal scanning method: L denotes lenses, D denotes photodetectors, and *z* is the scanning axis.

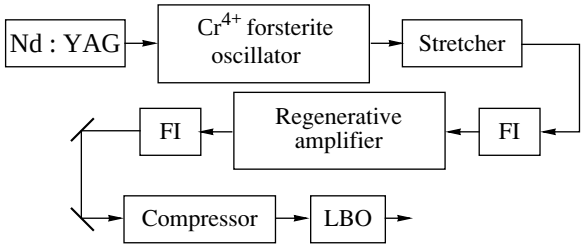

Fig. 4. Block diagram of a forsterite laser system for obtaining femtosecond radiation pulses.

mined from the expression for the total absorption coefficient α:

$$\alpha = \alpha_0 + \beta I, \tag{1}$$

where α_0 is the linear absorption coefficient and I is the radiation intensity. The nonlinear absorption coefficient β was measured by the longitudinal scanning method [22, 23, 25, 26]. The essence of this method is that transmission $T(z)$ is measured for the total luminous flux passing through the sample when it is displaced along the z axis of the focused laser beam (Fig. 3). In the vicinity of the focus ($z = 0$), the change in $T(z)$ associated with the nonlinear contribution to absorption is maximal; on the other hand, at a large distance from the focus, where the beam intensity is small, the medium behaves as a linear medium. The $T(z)$ dependence normalized to the transmission T_0 in the linear regime is defined by the relation [25, 26]

$$\frac{T(z)}{T_0} = \sum_{m=0}^{\infty} \frac{(-\beta I L_{\text{eff}})^m}{(m+1)^{3/2}[1 + (z/z_0)^2]^m}. \tag{2}$$

Here,

$$L_{\text{eff}} = \frac{1 - \exp(-\alpha_0 L)}{\alpha_0},$$

L is the film thickness, $z_0 = \pi\omega_0^2/\lambda$ is the confocal parameter, ω_0 is the radius of the laser beam waist, and λ is the wavelength.

To measure the nonlinear optical response of *J*-aggregates in the femtosecond range of pulse duration, we used second-harmonic generation of the forsterite laser as a source of exciting radiation since its wavelength coincides with the absorption peak of the films of *J*-aggregated under investigation. The block diagram of the femtosecond forsterite laser system is shown in Fig. 4. The setup included a master oscillator (pumped by a cw Nd : YAG laser manufactured by Spectra Physics Millennia IR), a stretcher, Faraday insulators (FI), a regenerative amplifier (pumped by a pulsed Nd : YAG Spectra Physics MERLIN laser), and a compressor. The main parameter of laser radiation at the computer exit were as follows: the pulse duration measured with the help of an autocorrelator was 75 fs at a pulse repetition rate $f = 1$ kHz and the central wavelength of generated pulses was 1250 nm for a spectral width of 26 nm. The radiation of the forsterite laser was doubled in the LBO crystal; as a result, we obtained radiation pulses with a wavelength of 625 nm for a pulse duration $t_p = 50$ fs. The waist diameter of the focused laser beam was 140 μm. In each position z, averaging was carried out over 75 pulses and the error in the measurement of $T(z)$ did not exceed 1–2%.

To study the nonlinear optical response of *J*-aggregates in the nanosecond range of pulse durations, we used a DCM dye laser pumped by the second harmonic of the Nd : YAG laser LTI-411 (pulse duration 5 ns, $f =$ 10 Hz, and the wavelength was tuned in the range 610–645 nm). The waist diameter of the focused laser beam was 100 μm. In each position z, averaging was carried out over 30 pulses and the error in the measurement of $T(z)$ at the chosen point on the film did not exceed 3–5%.

The absolute error in the measurement of the nonlinear response coefficient β, which is mainly determined by the error of measurements of the incident radiation intensity and the film thickness, was 25% both for femtosecond and for nanosecond pulse durations. The deviation in the values of β measured in different regions of the film was 20–30%. With increasing intensity of incident radiation, the change in the transmission $\Delta T = T(z) - T_0$ increased in the same proportion, indicating the correctness of measurements of the third-order nonlinear optical quantity. During the time of measurements of the $T(z)$ curve, a noticeable degradation in the film of *J*-aggregates as a result of laser radiation was not observed. All optical and nonlinear optical measurements were made at room temperature.

3. EXPERIMENTAL RESULTS AND DISCUSSION

Figure 5 shows typical longitudinal scanning curves obtained for different samples and for various pulse durations of exciting radiation. The curves in Fig. 5a for the normalized transmission T/T_0 of the dye film deposited on a substrate as a function of the sample position

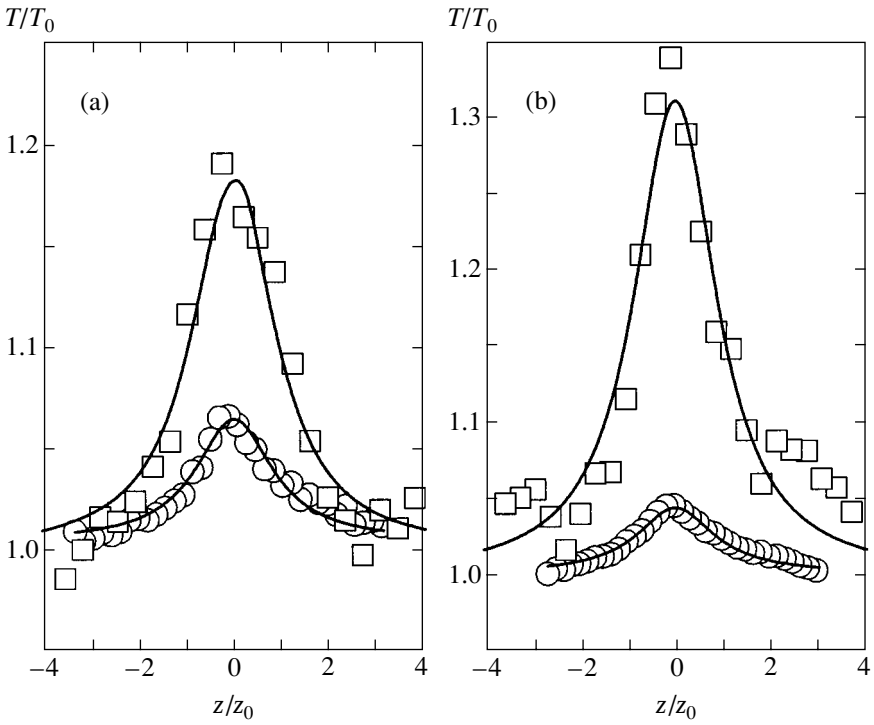

Fig. 5. Normalized transmission of the films of *J*-aggregates of thiacarbocyanine, measured at a wavelength $\lambda = 625$ nm. The radiation pulse duration is 5 ns (squares) and 50 fs (circles); solid curves correspond to fitting by the least squares method with the help of formula (2). (a) Film of *J*-aggregates with a dye : polymer weight ratio of 1 : 0, $L_{\text{eff}} = 18$ nm; fitting of circles gives $\beta = -1.5 \times 10^{-5}$ cm/W for $I = 6.2$ GW/cm^2; fitting of squares gives $\beta = -1$ cm/W for $I = 0.3$ MW/cm^2. (b) Polymer film containing *J*-aggregates with the dye : polymer weight ratio of 1 : 10, $L_{\text{eff}} = 466$ nm; fitting of circles gives $\beta = -7.5 \times 10^{-7}$ cm/W for $I = 3.5$ GW/cm^2; fitting of squares gives $\beta = -0.055$ cm/W for $I = 0.36$ MW/cm^2.

relative to the focal plane were measured under nanosecond excitation with a radiation wavelength $\lambda = 625$ nm (squares) and under femtosecond excitation by the second harmonic radiation from a forsterite laser (circles). The form of the measured $T(z)/T_0$ dependences corresponds to the bleaching of the exciton absorption peak of *J*-aggregates both for nanosecond and femtosecond pulse duration. It can be seen from Fig. 5a that the maximal bleaching of the film of *J*-aggregates at $z = 0$ under the action of nanosecond pulses corresponds to $\Delta T = 20\%$ for a radiation intensity $I = 0.3$ MW/cm^2 at the lens focus; at the same time, the value of ΔT for the same sample subjected to the action of femtosecond pulses with a radiation intensity $I = 6.2$ GW/cm^2 does not exceed 7–8%. The results of fitting (solid curves in Fig. 5a) the measured $T(z)/T_0$ dependences by the least squares technique with the help of formula (2) taking into account only the first two terms ($m \le 1$) with preset I and L_{eff} were used to determine the values of the nonlinear absorption coefficients β.

The typical value of β averaged over the measurements in various regions of the dye film deposited on a substrate was $-(1.8 \pm 0.3) \times 10^{-5}$ cm/W for femtosecond pulses and -1.3 ± 0.3 cm/W for nanosecond pulses. Thus, the values of β differ by a factor of $(7.2 \pm 2.0) \times 10^4$.

Analogous measurements for femto- and nanosecond pulses were made for a polymer film containing *J*-aggregates of thiacarbocyanine (see Fig. 5b). In this case, the nonlinear absorption coefficient was $\beta = -(1 \pm 0.2) \times 10^{-6}$ cm/W for femtosecond pulses and $\beta = -0.05 \pm 0.015$ cm/W for nanosecond pulses, which also leads to a difference in the measured values by a factor of $(5.0 \pm 1.8) \times 10^4$.

To analyze the observed difference in the absolute values of the nonlinear absorption coefficients for femto- and nanosecond pulses, we consider the time dependence of the saturation effect in the model of a two-level system interacting with resonance radiation. It should be noted that the reciprocal pulse duration $1/t_p$ for 50-fs pulses under our experimental conditions is substantially smaller (by a factor of 6–8) than the absorption line width of the molecular *J*-aggregates under investigation, so that the system dephasing over a pulse duration occurs quite rapidly and the populations change insignificantly. In such a case, we can disregard the evolution of the offdiagonal element in the density matrix (see, for example, [27]) and write the

balance of the populations for the ground (g) and the excited (e) states in the form

$$\frac{dN_e}{dt} = -A_{eg}N_e + Ib\Delta N, \quad (3)$$

$$\frac{dN_g}{dt} = A_{eg}N_e - Ib\Delta N. \quad (4)$$

Here, N_g and N_e are the populations of the corresponding states, $\Delta N = N_g - N_e$ is the difference in the populations, $N \equiv N_g + N_e$ is the concentration of particles, A_{eg} is the population relaxation constant for the upper level, I is the radiation intensity, and b is the probability of optical transitions in our system (the second Einstein coefficient). Over time intervals much smaller than the lifetime $\tau = 1/A_{eg}$, the rates of variation of the populations can be assumed to be constant and specified by the initial conditions $\Delta N(0) = N$. Then, the difference in the populations has the form

$$\Delta N(t) = N(1 - 2Ibt). \quad (5)$$

The nonlinear absorption coefficient β_p, which is a nonlinear response of the medium to the action of pulses of duration t_p, is proportional to the change in the population under the action of radiation:

$$\beta_p \propto -2Ibt_p. \quad (6)$$

In the case when the pulse duration is substantially larger than lifetime τ, we can assume that the problem is stationary. In this case, for the difference in the populations, we have

$$\Delta N = \frac{N}{1 + 2Ib\tau}. \quad (7)$$

Expanding this expression into a series in the intensity and retaining only the principal terms, we obtain

$$\Delta N \approx N(1 - 2Ib\tau). \quad (8)$$

As a result, we arrive at the stationary nonlinear absorption coefficient,

$$\beta_{st} \propto -2Ib\tau. \quad (9)$$

For nonlinear absorption coefficients measured for short pulses under steady-state conditions, we obtain the relation

$$\frac{\beta_{st}}{\beta_p} = \frac{\tau}{t_p}. \quad (10)$$

Using this relation and knowing the experimental data for β_{st}, β_p, and t_p, we can derive the value of τ.

The experimental results obtained by different research groups for various types of *J*-aggregates lead to lifetime $\tau \sim 10$–10^3 ps [6, 11, 19–21]. As compared to our experimental conditions, this value, on the one hand, substantially exceeds the pulse duration of 50 fs, which allows us to use for β_p and t_p the values corresponding to this regime of measurements. On the other, the value of the lifetime is much smaller than the pulse duration of 5 ns; consequently, the results of measurements under such conditions can be used as the value of β_{st}. As a result, for the experimentally measured values of β_{st}, β_p, and $t_p = 50$ fs, formula (10) gives the values of $\tau \approx 3.7 \pm 0.9$ ns for films without a polymeric matrix and $\tau \approx 2.9 \pm 0.5$ ns for films with a polymeric matrix, which contradict the available experimental data. Thus, the above analysis of the temporal nonlinear response of a simple two-level system shows that this system cannot be used for an adequate description of the observed nonlinear optical response of molecular *J*-aggregates and that additional relaxation processes should be taken into account.

Exciton–exciton annihilation is a well-known additional relaxation process. It should be noted, however, that, under our experimental conditions, the pulse duration of 50 fs is substantially smaller than the characteristic time constants for exciton–exciton annihilation processes, which are on the order of 1 ps [18, 21, 28, 29]; consequently, the influence of this effect is weak for the given regime of measurements. For measurements in the nanosecond range of pulse duration, allowance for the exciton–exciton annihilation would result in a decrease in the exciton lifetime and in the value of the nonlinear optical response [18, 21, 28, 29]; however, the effect observed by us here on the contrary leads to an increase in the nonlinear optical response for nanosecond pulses. Thus, the effect of exciton–exciton annihilation cannot provide a correct description of the observed behavior of the nonlinear optical response of molecular *J*-aggregates.

One of the most probable additional relaxation channels, which may lead to the observed behavior of the nonlinear optical response, is the formation of self-trapped exciton states [30]. The emergence of exciton self-trapping at low temperatures was observed in molecular *J*-aggregates of pseudoisocyanine [31–33] and in aggregates of TDC dyes [34] in the form of a fluorescence peak displaced towards the low-frequency spectral region. The dynamics of self-trapped states was studied in [31, 32], where it was found that the Stokes shift of the fluorescence peak increases with time and the decay of fluorescence in the vicinity of the *J* peak is nonmonoexponential. An analogous fluorescence peak was also observed for *J*-aggregates of pseudoisocyanine at room temperatures [22], Finally, fluorescence in the range of 650–700 nm was observed under our experimental conditions during the excitation of films of *J*-aggregates of thiacarbocyanine dye at a wavelength $\lambda = 580$ nm. The dot-and-dash curves in Fig. 2 show the fluorescence spectra of films containing *J*-aggregates of thiacarbocyanine at room temperature.

It should be noted that, in these measurements, the same films as in the nonlinear optical measurements were used (the optical density of these films attained almost 1.0 optical unit). Consequently, fluorescence of free excitons in *J*-aggregates, which is characterized by a Stokes shift on the order of several nanometers, was strongly suppressed due to reabsorption. For this reason, we could apparently observe only the far fluorescence wing associated with exciton emissions from self-trapped states.

The exciton localization mechanism in molecular *J*-aggregates is the change in the local structure of a molecular aggregate upon the formation of an exciton, e.g., as a result of a change in the constant dipole moment of dye molecules forming the aggregate, under the action of optical excitation [23, 35]. In this case, the translational symmetry of the aggregate is broken and, hence, an exciton cannot propagate along the aggregate. Thus, the exciton is self-trapped in a certain region of the molecular aggregate. Since the energy of the exciton state is mainly determined by the energy of interaction of neighboring molecules in the aggregate, the energy of interaction will also change as a result of the change in the structure, which shifts the self-trapped exciton state to a position below the bottom of the exciton band. In addition, it is important that the lifetime of the localized exciton state is longer than that of a delocalized state (this was observed in [31, 32]). The increase in the lifetime is a consequence of its proportionality to the length of the ordered segment of the aggregate, on which the exciton is localized [10]. In the experiments [31, 32], the lifetime of a localized exciton was estimated at 1–10 ns.

Let us construct a three-level model to analyze the nonlinear optical response of molecular *J*-aggregates taking into account the additional relaxation channel (Fig. 6). Here, g is the ground level, e is the excited level of free excitons, a is the level of self-trapped excitons, and A are the relaxation constants for the corresponding decay channels; the A_{ea} relaxation is nonradiative and radiation at frequency ω is in resonance with the e–g transition.

The system of balance equations for this model has the form

$$\frac{dN_e}{dt} = -(A_{eg} + A_{ea})N_e + Ib\Delta N, \tag{11}$$

$$\frac{dN_g}{dt} = A_{ag}N_a + A_{eg}N_e - Ib\Delta N, \tag{12}$$

$$\frac{dN_a}{dt} = -A_{ag}N_a + A_{eg}N_e, \tag{13}$$

where $\Delta N = N_g - N_e$ is the difference in the populations and $N \equiv N_g + N_e + N_a$ is the concentration of particles. Analogously to the case of a two-level system, we find the difference in the populations for short time intervals:

$$\Delta N(t) = N(1 - 2Ibt). \tag{14}$$

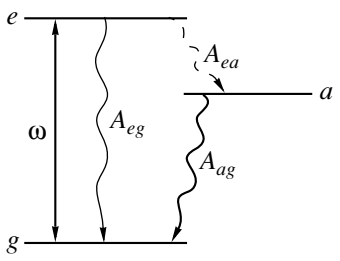

Fig. 6. Three-level model constructed for describing the nonlinear optical response of molecular *J*-aggregates.

In this case, the evolution of the system is determined by the initial conditions under which only the ground state is populated and radiation-induced transitions to additional level a are absent; consequently, the solution coincides with that for the two-level system. For long pulses, we solve the stationary problem, which leads to the following expression for the difference in the populations in the three-level system:

$$\Delta N = \frac{N}{1 + 2Ib\tau_t} \approx N(1 - 2Ib\tau_t), \tag{15}$$

in this case, the relation for the nonlinear absorption coefficients for short and long pulses assumes the form

$$\frac{\beta_{st}}{\beta_p} = \frac{\tau_t}{t_p}. \tag{16}$$

Here, $\tau_t = 1/A_t$, where A_t is defined as

$$A_t = 2A_{ag}\frac{A_{ea} + A_{eg}}{A_{ea} + 2A_{ag}}. \tag{17}$$

If the excited level e does not decay to the additional level a ($A_{ea} = 0$), we have $A_t = A_{eg}$, and the system behaves as a two-level system with the corresponding relaxation rate. Otherwise, when relaxation from level e to level a is rapid as compared to the remaining decay channels ($A_{ea} \gg A_{eg}, A_{ag}$), the nonlinear optical response of the system is completely determined by the relaxation constant of the additional level $A_t = 2A_{ag}$. As was mentioned above, the relaxation of the self-trapped state (additional level) for molecular aggregates is slower than the relaxation of the free exciton level (excited level). In this case, the behavior of the system reflects the well-known optical pumping effect. As a result of accumulation of population on the additional level, this effect leads to a substantial enhancement of bleaching of the medium for long pulses or cw radiation, which is precisely observed in our experiments.

Thus, the value of $\tau_t \sim 3.5$ ns obtained above from a comparison of our experimental data for the nonlinear absorption coefficients of *J*-aggregates for short (50 fs) and long (5 ns) pulses is characterized not by the lifetime τ of the free exciton state, which is substantially smaller than this value, but by a certain "effective" lifetime of an self-trapped exciton (see formula (17)). A more detailed description of the observed effect requires experimental data on the quantities A_{eg}, A_{ea}, and A_{ag}, which is the subject of our subsequent studies.

4. CONCLUSIONS

J-aggregates of thiacarbocyanine dye with the exciton absorption peak in the region of 630 nm, which coincides with the wavelength of second harmonic generation of a femtosecond forsterite laser, have been obtained. The measurements of the nonlinear absorption coefficient performed by longitudinal scanning using femtosecond and nanosecond pulses of exciting radiation revealed an increase in the nonlinear absorption coefficient with increasing pulse duration. The analysis of the results on the basis of two- and three-level models indicates that allowance for an additional level emerging due to the formation of self-trapped exciton states leads to good qualitative agreement with the observed effect. Accumulation of excitons in self-trapped states leads to a substantial enhancement of bleaching of the medium under the action of nanosecond laser pulses.

ACKNOWLEDGMENTS

This study was supported financially by the Russian Foundation for Basic Research (project no. 02-03-33336) and the Complex Program for Fundamental Studies of the Presidium of the Russian Academy of Sciences (project no. 8–2).

REFERENCES

1. *J-aggregates*, Ed. by T. Kobayashi (World Sci., Singapore, 1996).
2. E. W. Knapp, Chem. Phys. **85**, 73 (1984).
3. H. Fidder, J. Knoester, and D. A. Wiersma, J. Chem. Phys. **95**, 7880 (1991).
4. F. C. Spano and S. Mukamel, Phys. Rev. A **40**, 5783 (1989).
5. Y. Wang, J. Opt. Soc. Am. B **8**, 981 (1991).
6. V. L. Bogdanov, E. N. Viktorova, S. V. Kulya, and A. S. Spiro, Pis'ma Zh. Éksp. Teor. Fiz. **53**, 100 (1991) [JETP Lett. **53**, 105 (1991)].
7. F. A. Zhuravlev, N. A. Orlova, V. V. Shelkovnikov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **56**, 264 (1992) [JETP Lett. **56**, 260 (1992)].
8. J. Knoester, Chem. Phys. Lett. **203**, 371 (1993).
9. R. V. Markov, A. I. Plekhanov, V. V. Shelkovnikov, and J. Knoester, Microelectron. Eng. **69**, 528 (2003).
10. H. Fidder, J. Knoester, and D. A. Wiersma, Chem. Phys. Lett. **171**, 529 (1990).
11. M. Furuki, M. Tian, Y. Sato, *et al.*, Appl. Phys. Lett. **77**, 472 (2000).
12. V. I. Avdeeva, A. S. Kuch'yanov, A. I. Plekhanov, *et al.*, Kvantovaya Élektron. (Moscow) **33**, 539 (2003).
13. A. S. Kuch'yanov, R. V. Markov, A. I. Plekhanov, *et al.*, Opt. Commun. **231**, 343 (2004).
14. O. Kuhn and V. Sundstrom, J. Chem. Phys. **107**, 4154 (1997).
15. M. Yang, R. Agarwal, and G. R. Fleming, J. Photochem. Photobiol. A **142**, 107 (2001).
16. V. Malyshev and P. Moreno, Phys. Rev. A **53**, 416 (1996).
17. E. I. Malt'sev, D. A. Lypenko, B. I. Shapiro, *et al.*, Appl. Phys. Lett. **75**, 1896 (1999).
18. K. Minoshima, M. Taiji, K. Misawa, and T. Kobayashi, Chem. Phys. Lett. **218**, 67 (1994).
19. K. Ohta, M. Yang, and G. R. Fleming, J. Chem. Phys. **115**, 7609 (2001).
20. H. Fidder, J. Knoester, and D. A. Wiersma, J. Chem. Phys. **98**, 6564 (1993).
21. R. Gadonas, K.-H. Feller, and A. Pugzlys, Opt. Commun. **112**, 157 (1994).
22. R. V. Markov, A. I. Plekhanov, S. G. Rautian, *et al.*, Opt. Spektrosk. **85**, 643 (1998) [Opt. Spectrosc. **85**, 588 (1998)].
23. R. V. Markov, A. I. Plekhanov, V. V. Shelkovnikov, and J. Knoester, Phys. Status Solidi B **221**, 529 (2000).
24. N. A. Orlova, E. F. Kolchina, M. M. Shakirov, *et al.*, Zh. Org. Khim. **74**, 256 (2004).
25. M. Sheik-bahae, A. A. Said, and E. W. Van Stryland, Opt. Lett. **14**, 955 (1989).
26. M. Sheik-bahae, A. A. Said, T.-H. Wei, *et al.*, IEEE J. Quantum Electron. **26**, 760 (1990).
27. L. Allen and J. H. Eberly, *Optical Resonance and Two-Level Atoms* (Wiley, New York, 1975; Mir, Moscow, 1978).
28. V. Sundstrom, T. Gillbro, R. A. Gadonas, and A. Piskarskas, J. Chem. Phys. **89**, 2754 (1988).
29. E. Gaizauskas, K.-H. Feller, and R. Gadonas, Opt. Commun. **118**, 360 (1995).
30. E. I. Rashba, in *Excitons*, Ed. by E. I. Rashba and M. D. Sturge (North-Holland, Amsterdam, 1982; Nauka, Moscow, 1985).
31. Yu. V. Malyukin and O. G. Tovmachenko, Pis'ma Zh. Éksp. Teor. Fiz. **58**, 358 (1993) [JETP Lett. **58**, 393 (1993)].
32. Yu. V. Malyukin, V. P. Seminozhenko, and O. G. Tovmachenko, Zh. Éksp. Teor. Fiz. **107**, 812 (1995) [JETP **80**, 460 (1995)].
33. G. S. Katrich, K. Kemnitz, Yu. V. Malyukin, and A. M. Ratner, J. Lumin. **90**, 55 (2000).
34. M. A. Drobizhev, M. N. Sapozhnikov, I. G. Scheblykin, *et al.*, Chem. Phys. **211**, 455 (1996).
35. K. Misawa and K. Minoshima, Chem. Phys. Lett. **220**, 251 (1994).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 99, No. 3, 2004, pp. 487–493.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 3, 2004, pp. 558–565.

PLASMA, GASES

Generation of Harmonics during the Interaction of Ultrashort Superstrong Laser Pulses with Solid Targets

V. S. Rastunkov and V. P. Krainov*
Moscow Institute of Physics and Technology, Dolgoprudnyĭ, Moscow oblast, 141700 Russia
**e-mail: krainov@online.ru*
Received March 22, 2004

Abstract—Generation of even and odd harmonics in the skin layer formed during the interaction of a short relativistic laser pulse with solid targets is considered. The complex motion of free electrons in the skin layer along the electric field vector and along the direction of propagation of a laser wave is analyzed. The Fourier expansion of the trajectory of this motion is used to obtain the components of the conductivity tensor and of the amplitude of the transverse electromagnetic field of harmonics propagating along the electric field. Even harmonics appear due to relativistic effects. The efficiency of generation of even and odd harmonics at the leading front of a laser pulse is calculated. *© 2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The action of strong laser fields on a dense or a rarefied medium induces a fundamental nonlinear process: generation of higher harmonics of the laser field. In particular, when laser radiation interacts with solid targets, a preplasma and a overdense plasma consisting of multiply charged atomic ions and electrons are formed. Our task is to study induced emission of free electrons from solid targets in extremely strong laser fields. Two mechanisms of harmonic generation can be singled out, namely, the emission from bound and free electrons. These two possibilities are analyzed, for example, in [1] in the case of harmonic generation by gases (underdense plasma) for various substances. The efficiency of even harmonic generation is much lower than for odd harmonics in gaseous He, while these efficiencies are comparable in gaseous N_2 and Ar. Harmonic generation in a rarefied (subcritical plasma) was also studied theoretically in [2, 3].

The interaction of laser radiation with a dense medium should be considered separately. It is well known that a high efficiency of transformation of the laser energy into higher harmonic generation can be attained when a laser pulse propagates in a hollow plasma channel. Even and odd harmonics are then generated when radiation is reflected by channel walls. The corresponding 2D problem was considered in [4]. A similar 1D problem was analyzed theoretically in [5], where the possibility of harmonic generation was demonstrated. In the context of these problems, we proposed the solution to the problem of the behavior of an electron in a overdense plasma in the field of a relativistic laser pulse [6]. In the present paper, the results obtained in [6] are used to consider generation of (primarily even) harmonics in a overdense plasma.

When ultrastrong laser fields act on a solid target, a skin layer formed near the surface prevents the penetration of the electric field in the substance to a considerable extent. Here, we do not consider the ionization mechanism (tunnelling, above-the-barrier, or collisional) of formation of free electrons and atomic ions. It will be shown below that allowance for even a weak electric field in the skin layer plays a significant role. The motion of charges becomes nonlinear, which leads to harmonic generation in the skin layer. Naturally, with such a formulation of the problem we consider only bulk effects (pertaining to thickness of the substance on the order of the skin depth). Moreover, we consider harmonic generation by free electrons only during their collisions with atomic ions.

The problem is solved according to the procedure used earlier in [2, 3]. In Section 2, the equations of motion of an electron in a substance are analyzed analytically and numerically. For convenience of subsequent integration, the numerical results are approximated by analytic dependences with parameters determined from numerical calculations. The results are compared to those obtained using an analytic approach, which gives an asymptotic behavior of momenta for moving particles. In Section 3, using the momenta as functions of time, we derive an expression for the electric current component along the polarization of the laser field. We do not consider longitudinal generated electric fields since these fields do not leave the plasma volume [3]. Expanding the expression for the current density into a Fourier series, we obtain the conductivity tensor components. In Section 4, the electric field components at frequencies of harmonics are analyzed and an expression for the generation efficiency is obtained.

As was shown in [6], electrons leaving the field region possess a drift momentum along the direction of

propagation of the external field. An electron emitted from the skin layer has a nonzero drift momentum. This is mainly due to the fact that the electric and magnetic components in the skin layer are shifted in phase by $\pi/2$ relative to one another. The electron velocity is quite high, though not relativistic. An electron beam moving with this velocity induces annular magnetic field of a huge intensity [7]. According to the results of our preliminary calculations, the magnetic pressure produced by such fields cannot ensure the electron beam stability.

2. NONLINEAR MOTION OF AN ELECTRON IN THE SKIN LAYER UNDER THE ACTION OF THE LASER PULSE FIELD

We assume for simplicity that a linearly polarized laser beam propagates at right angles to the surface of a dense medium. We direct the x axis along the direction of propagation of the laser pulse, the y axis, along the electric field vector, and the z axis, along the magnetic field. The relativistic plasma frequency is given by [8]

$$\omega_p = \sqrt{4\pi n_e/\gamma},$$

where n_e is the electron concentration in the skin layer (here and below, we use the relativistic system of units, in which $c = e = m = 1$); γ is the relativistic factor,

$$\gamma = \sqrt{1 + (F^{in}/\omega)^2} > 1,$$

and F^{in} is the amplitude of the electric field in the plasma. We assume that $\omega_p \gg \omega$ (this condition is usually satisfied when solids are irradiated by ultrastrong laser pulses on account of multiple ionization of atoms). Since the electric field in the skin layer is weak as compared to the electric field of the incident wave, we can use the nonrelativistic value of the plasma frequency ($\gamma \approx 1$) up to an incident radiation intensity on the order of 10^{21} W/cm^2.

The boundary conditions at the vacuum-overdense plasma interface have the familiar form

$$F^{in} = \frac{2}{1+\sqrt{\varepsilon}}F(t), \quad B^{in} = \frac{2\sqrt{\varepsilon}}{1+\sqrt{\varepsilon}}F(t).$$

Here, $F(t)$ is the amplitude of the electric field of the electromagnetic wave in vacuum, ω is its frequency, B^{in} is the amplitude of the magnetic field strength in the plasma,

$$\varepsilon = 1 - \frac{\omega_p^2}{\omega^2}$$

is the dielectric constant of the overdense plasma ($|\varepsilon| \gg 1$) produced by free electrons. The laser pulse is turned on and off adiabatically and is approximated by the Gaussian envelope

$$F(t) = F\exp(-t^2/\tau^2),$$

where τ is the laser pulse duration and the condition $\omega\tau \gg 1$ is satisfied.

The classical Newton equations for relativistic pulses p_x and p_y within the skin layer ($x > 0$) have the form (the x axis is directed along the normal to the target surface, while the y axis is directed along the electric field of a linearly polarized laser pulse)

$$\frac{dp_x}{dt} = 2v_yF(t)\exp(-x/\delta)\sin(\omega t), \tag{1}$$

$$\frac{dp_y}{dt} = 2\frac{\omega}{\omega_p}F(t)\exp(-x/\delta)\cos(\omega t) - 2v_xF(t)\exp(-x/\delta)\sin(\omega t), \tag{2}$$

where $\delta = 1/\omega_p$ is the skin depth and v_x and v_y are the electron velocity components, i.e.,

$$v_x = \frac{p_x}{\sqrt{1+p_x^2+p_y^2}}, \quad v_y = \frac{p_y}{\sqrt{1+p_x^2+p_y^2}}. \tag{3}$$

The value of $x = 0$ corresponds to the target surface. We disregard the motion of the electron along the z axis. The results are virtually independent of the initial phase of the laser wave since the laser pulse is actuated adiabatically.

It will be shown below that an electron in the skin layer possesses a relatively high drift velocity along the x axis and a comparable vibrational velocity along the y axis. Consequently, the disregard of electron collisions with atomic ions is justified. Equations (1)–(3) can be solved with the following initial conditions: in the limit $t = -\infty$, an electron is at rest at point $x = y = 0$.

We will first find the solution in the approximation of weak relativism for the time interval $t \in (-\infty, t_0)$, where $|t_0| \gg \tau$. Under these assumptions, we can set for simplicity $\exp(-x/\delta) \approx 1$. In addition, we can disregard the second term in Eq. (2), corresponding to the magnetic component of the Lorentz force as compared to the first term corresponding to the electric component. We can now integrate Eq. (2) approximately, bearing in mind that $F(t)$ is a more slowly varying function as compared to $\cos(\omega t)$. Thus, we obtain

$$v_y(t) \approx \frac{2}{\omega_p}F(t)\sin(\omega t). \tag{4}$$

Substituting this result into Eq. (1), we find that

$$v_x \approx -\frac{F_0^2\tau^2}{2\omega_p t}\exp\left(-\frac{2t^2}{\tau^2}\right). \tag{5}$$

This expression represents the relation between the final velocity and the sought time during which the

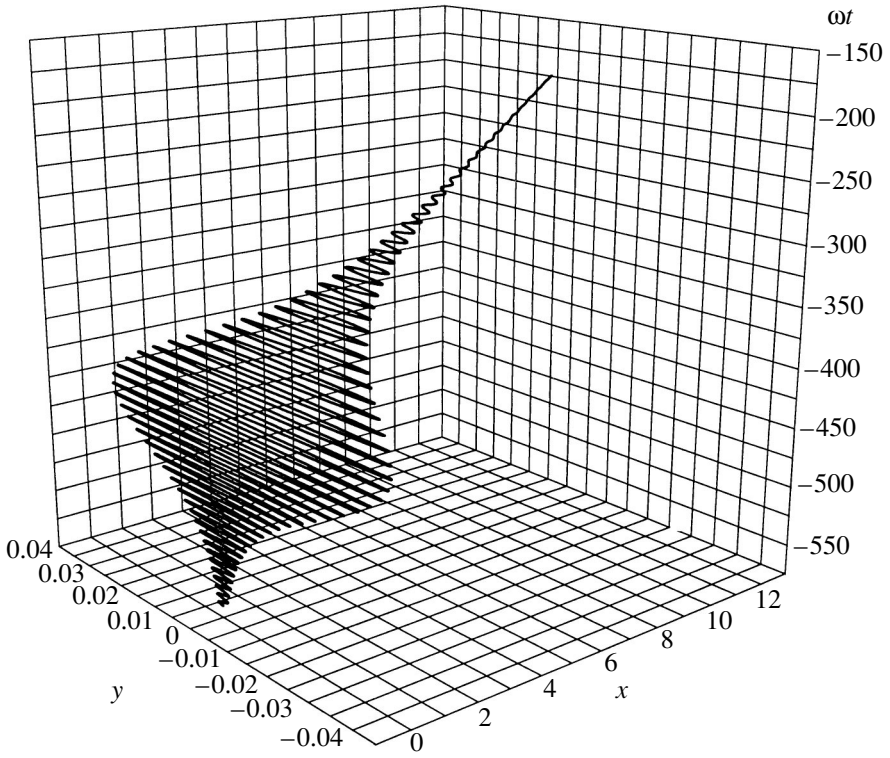

Fig. 1. Typical 2D trajectory of an electron in the skin layer along the direction of the electric field of a laser wave (*y* axis) and along the direction of propagation of the laser wave (*x* axis) at the leading front of the laser pulse. Both electron coordinates are given in units of the skin depth as functions of the field phase ω*t*.

electron is in the field region (within the skin layer). Let us estimate this time. We have

$$\delta = \int_{-\infty}^{t} v_x dt \overset{|t| \gg \tau}{\approx} \frac{F_0^2 \tau^2}{2\omega_p t} \frac{\tau^2}{4t} \exp\left(-\frac{2t^2}{\tau^2}\right). \quad (6)$$

Taking into account the fact that the skin depth $\delta = 1/\omega_p$, we obtain the estimate

$$t \approx \tau \sqrt{\ln\left(\frac{F_0 \tau}{8}\right)}. \quad (7)$$

If the intensity of incident radiation is $I = 5 \times 10^{19}$ W/cm^2, the amplitude of the electric field is $F_0 \approx 1.91 \times 10^{11}$ V/cm. We assume that the pulse duration τ is 83 fs. For the required time, we obtain $|\omega t| \approx 412$. This leads to the conclusion that an electron escapes from the skin layer with a nonzero drift momentum along the *x* axis. This is mainly due to the fact that the electric and magnetic field components in the skin layer are shifted in phase by π/2 relative to each other (this is so if $\omega_p > \omega$).

The motion of an electron is determined by two dimensionless parameters $f = 2F/\omega$ and ω_p/ω, we choose the characteristic values of these parameters: the photon energy $\hbar\omega = 1.5$ eV, the frequency ratio $\omega_p/\omega = 7$, and $f = 10$ for an amplitude value of intensity of 5×10^{19} W/cm^2.

The results of numerical calculations for these parameters are given in our previous publication [6]. Figure 1 shows the 2D trajectory of an electron. This solution and be approximated analytically as follows:

$$p_x(t) = \frac{a_1}{1 + \exp(-a_2(\omega t - a_3))},$$
$$p_y(t) = a_4 \exp\left(-\left(\frac{\omega t - a_5}{a_6}\right)^2\right)\sin(\omega t). \quad (8)$$

The parameters have the following numerical values:

$$a_1 = 0.0068, \quad a_2 = 0.0538, \quad a_3 = -422.640,$$
$$a_4 = 0.00486, \quad a_5 = -422.640, \quad a_6 = 65.4.$$

Figure 2 shows for comparison the results of numerical and analytical calculations. The discrepancy is for the following two reasons. First, the particle velocity becomes quite large and the weak relativistic approximation cannot be used. Second, the particle penetrates the unperturbed region of the plasma under the skin layer and, hence, the factor $\exp(-x/\delta)$ in Eqs. (1) and (2) cannot be neglected.

It should be emphasized that, in the case a underdense plasma (or vacuum), the electron drift along the

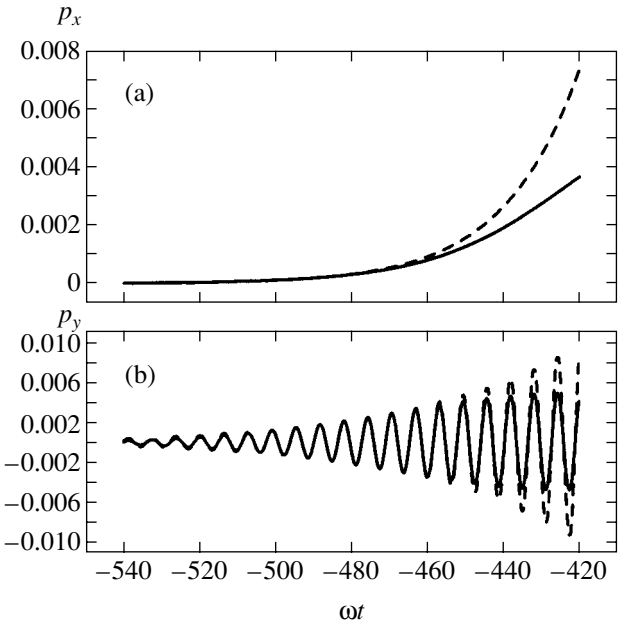

Fig. 2. Electron momenta along the direction of the electric field of a laser wave (*y* axis) and along the direction of propagation of the laser wave (*x* axis) during the motion of the electron in the skin layer, shown in Fig. 1. The values of momenta are given in units of *mc* (solid curves correspond to the results of numerical calculations and dashed curves are calculated analytically).

direction of propagation of a femtosecond laser pulse is determined exclusively by the laser pulse envelope. The relativistic motion of the an electron along a trajectory resembling Fig. 8 becomes unclosed due to an increase or decrease in the wave amplitude during the laser pulse action. On the contrary, the electron drift in a overdense plasma is not directly connected with the laser pulse envelope.

3. CONDUCTIVITY OF THE MEDIUM AT THE FREQUENCY OF HARMONICS

When an electrons collides with an atomic ion having a charge *Z*, the transport cross section of elastic relativistic scattering through small angles is defined by the Mott formula [9] (in relativistic units)

$$\sigma_M = \frac{4\pi Z^2 \Lambda}{p^2(t) v^2(t)}. \quad (9)$$

Here, Λ is the Coulomb logarithm and $p(t)$ and $v(t)$ are the electron momentum and velocity. In the high-velocity limit, the Coulomb logarithm is quantum-mechanical [10].

The frequency of elastic electron-ion collisions is given by

$$\nu_{ei} = \sigma_M N_i v = \frac{4\pi Z^2 N_i \Lambda}{p^2(t) v(t)}. \quad (10)$$

Here, N_i is the concentration of atomic ions. Multiplying this expression by the electron velocity vector **v**, the electron concentration N_e, and the time interval dt, we obtain the following expression for the electron current density [11–13]:

$$d\mathbf{j} = -N_e \mathbf{v} \nu_{ei} dt. \quad (11)$$

This quantity has components along the *x* and *y* axes. It should be noted that this relation is also valid in the relativistic case (the so-called Pauli formula [14]).

The *x* component of current (11) along the *x* axis is responsible for the longitudinal electric field, which is absent outside the plasma [13]. For this reason, we will henceforth concentrate our attention only at the *y* component of the current density. Substituting the expressions obtained above for the total electron velocity and momentum into Eq. (11), we obtain

$$dj_y = -AFf(t)dt. \quad (12)$$

Here, the notation

$$A = \frac{4\pi Z^2 N_e N_i \Lambda}{F} \quad (13)$$

is introduced and the function

$$f(t) = \frac{p_y(t)}{(p_x^2(t) + p_y^2(t))^{3/2}} \quad (14)$$

is defined.

Integrating Eq. (12) and expanding function $f(t)$ into a Fourier series, we obtain the current density component along the *y* axis:

$$j_y = -AFa_0\omega t - AF\sum_{k=1}^{\infty}\left(\frac{a_k}{k}\sin(k\omega t) - \frac{b_k}{k}\cos(k\omega t)\right),$$

$$\omega t \in [s\pi, s\pi + 2\pi], \quad s \in \mathbb{Z}, \quad k = 1, 2, \ldots.$$

Here, a_0, a_k, and b_k are the Fourier coefficients. This leads to the following expression for the conductivity tensor:

$$\sigma_y^{(n)} = -A\frac{\sqrt{a_n^2 + b_n^2}}{n}\sin\left(n\omega t - \arctan\left(\frac{b_n}{a_n}\right)\right). \quad (15)$$

Function (14) is not periodic; consequently, we must analyze expansion (15) for various values of phase πs. Figure 3 shows four graphs representing the values of the relative amplitudes of the conductivity tensor of harmonics at different instants at the leading front of a laser pulse. The relative value of harmonics decreases with time during the buildup of the laser pulse. This is explained by the escape of an electron from the skin region to the unperturbed region of the substance. The

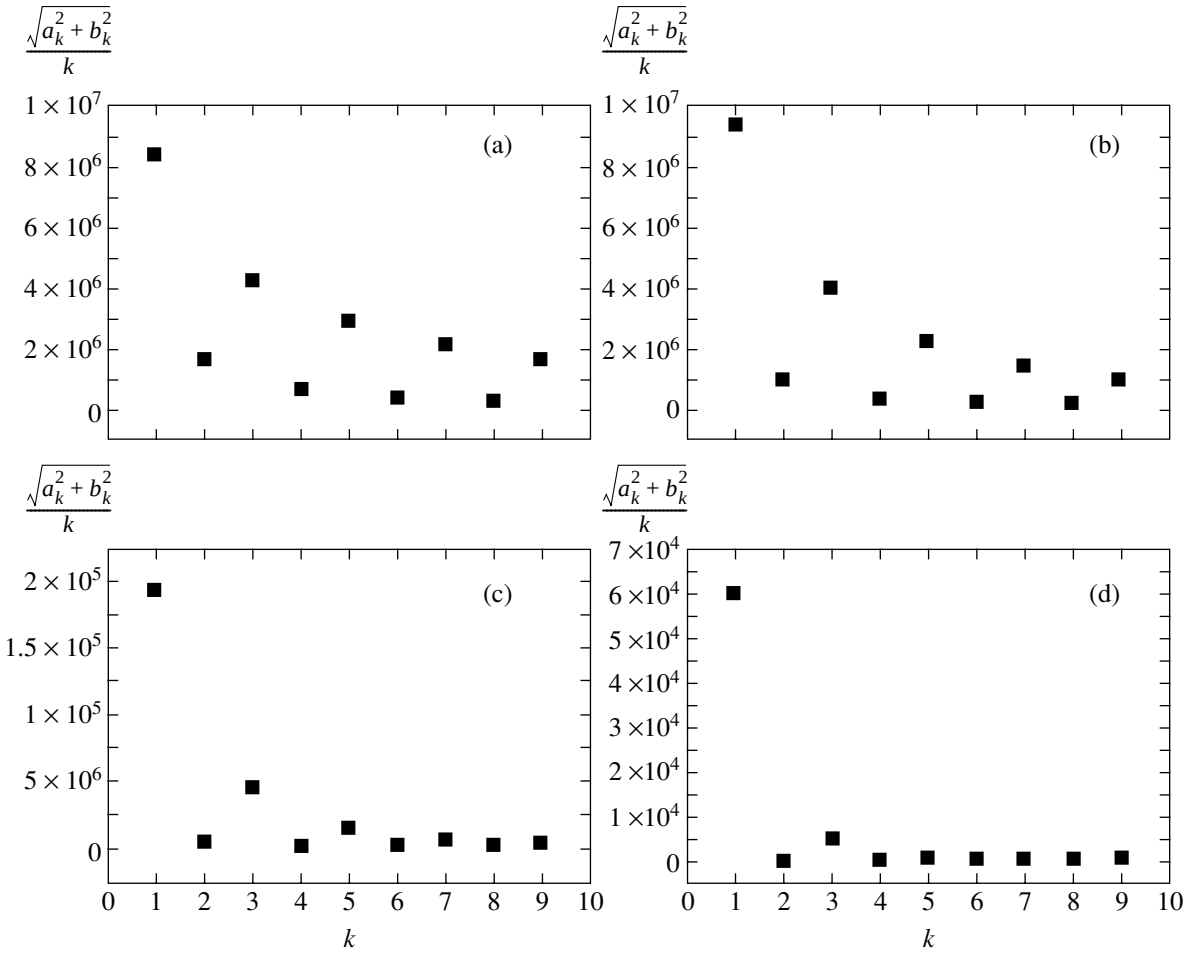

Fig. 3. Relative amplitudes of the conductivity tensor of harmonics at the instants characterized by phase ωt at the leading front of a laser pulse as function of the harmonic number k; $\omega t = -160\pi$ (a), -150π (b), -140π (c), and -130π (d).

efficiency of even harmonic generation is much lower than the efficiency of generation of odd harmonics.

4. INTENSITY OF RELATIVISTIC HARMONICS

Proceeding in the same way as in [3], we can write the equation for the electric field strength at the frequency of the nth harmonic:

$$-\frac{\partial^2 F_y^{(n)}}{\partial t^2} + c^2\frac{\partial^2 F_y^{(n)}}{\partial x^2} - 4\pi N_e F_y^{(n)} = AF\omega c\sqrt{a_n^2 + b_n^2}\cos\left(n\omega t - \arctan\frac{b_n}{a_n}\right). \quad (16)$$

The solution to this equation has the form

$$F_y^{(n)} = \frac{AFc\sqrt{a_n^2 + b_n^2}}{\omega\left(n^2 - \frac{\omega_p^2}{\omega^2}\right)}\cos\left(n\omega t - \arctan\frac{b_n}{a_n}\right), \quad (17)$$

where $\omega_p = \sqrt{4\pi N_e}$. For the ratio of intensity of the harmonic to the intensity of the external magnetic field, we obtain from Eq. (17) the analytic expression

$$\eta^{(n)} = \frac{|F_y^{(n)}|^2}{|F\cos(\omega t)|^2} = \left|\frac{4\pi Z^2 N_e N_i \Lambda c\sqrt{a_n^2 + b_n^2}}{F\omega\left(n^2 - \omega_p^2/\omega^2\right)}\right|^2. \quad (18)$$

It should be noted that the relative efficiency of harmonic generation (18) decreases both with increasing number of harmonic and with increasing external field intensity. It was shown above that the efficiency also decreases with time (coefficients a_n and b_n decrease).

5. PROPAGATION OF AN ELECTRON BEAM IN A OVERDENSE PLASMA

For the electron beam radius, we take the typical value r = 5–10 μm, while for the free electron number density, we obtain $n_e = 5 \times 10^{22}$ cm^{-3}. It is well known

that the expression for the magnetic pressure can be written in the form

$$p_M = \frac{H^2}{8\pi}.$$

The azimuthal magnetic field at distance r from the pulse axis is given by

$$H_\phi = \frac{2J(r)}{cr}.$$

To obtain estimates, we consider the limiting case of an ideal gas with the equation of state $p = n_e k_B T$. Let us estimate the temperature at which the z-pinch effect is in equilibrium,

$$J \sim n_e e u \pi r^2,$$

whence

$$H_\phi = \frac{2n_e e u \pi r}{c}.$$

The numerical value of H_ϕ is approximately 7500 T. We have

$$\frac{H_\phi^2}{8\pi} = n_e k_B T,$$

which gives

$$T = \frac{1}{n_e k_B}\frac{H_\phi^2}{8\pi} \approx 2.8 \times 10^3 \text{ eV}.$$

Let us estimate the temperature proceeding from the value of the mean energy of the gas. For the temperature corresponding to the drift motion along the x axis, we obtain the estimate

$$T = \frac{m_e u^2}{k_B} \approx 0.51 \text{ eV},$$

while for the temperature corresponding to the vibrations (quiver motion) along the y axis, we obtain

$$T = \frac{m_e v_{\text{quiver}}^2}{k_B}.$$

The following expression could serve as a natural estimate for the quiver pulse:

$$p_F \equiv p_{\text{quiver}} = eF/\omega.$$

However, in view of the weakness of the electric field in the skin layer, such an estimate will be substantially exaggerated. We will use the characteristic value of this momentum obtained in numerical experiment, which gives

$$T = 12.8 \text{ eV}.$$

Comparing the temperatures corresponding to the translational and quiver motion of an electron with the temperature at which the z pinch is in equilibrium, we could apparently conclude that the beam should contract. However, the requirement on plasma quasi-neutrality is important; if this requirement is violated, the electrons fly apart under the action of Coulomb forces. Let us estimate the velocity of atomic ions. Since

$$\frac{v_e}{v_i} \sim \sqrt{\frac{M}{m_e}},$$

we obtain

$$v_i \sim v_e \sqrt{\frac{m_e}{M}}.$$

It can be seen that in the general case the velocity of ions is approximately two orders of magnitude lower than the electron velocity. When the regime of ambipolar diffusion is realized, both types of charged particles move with close velocities (provided that the electron mobility is much higher than the ion mobility). Let us estimate the characteristic time of ion emission from the skin layer:

$$\tau_i = \frac{\delta}{v_i} \approx \frac{c}{\omega_p v_e}\sqrt{\frac{M}{m_e}} \approx 6 \times 10^{-12} \text{ s}.$$

This time is two orders of magnitude larger than the corresponding value for electrons,

$$\tau_e = \frac{\delta}{v_e} \approx \frac{c}{\omega_p v_e} \approx 6 \times 10^{-14} \text{ s}.$$

It is well known [15] that the quasi-neutrality transient time is

$$\tau_0 \sim \frac{1}{4\pi^2 N_e K_e}. \quad (19)$$

Here, K_e is the electron mobility. Assuming for estimates that $K_e \sim 10^5$ cm^2/(V s) (although this value is exaggerated), we obtain $\tau_0 \sim 10^{-7}$ s. According to this estimate, beam pinching cannot be observed in this case, and faster electrons leave the field region and fly apart, losing their energy as a result of collisions (outside the field region, the vibrational velocity is zero and collisions cannot be discarded). This is due to the relatively long quasi-neutrality transient time and the small skin depth.

Electron beam pinching was recently observed in the experiments [16] on the Vulcan laser facility for an intensity of 5×10^{19} W/cm^2 of a laser beam interacting with thin wires. Beg *et al.* [16] noted an interesting effect, i.e., the existence of returned electron current in the cold plasma surrounding the electron beam. The physical reason for this phenomenon is as follows. If an electron beam propagates in vacuum, the ring-shaped magnetic field slowly decreases with increasing distance for the beam axis. If, however, a cold plasma exists around the electron beam, the magnetic field penetrates only in a small region outside the beam (skin layer). It is the electron returned current emerging in

the cold plasma that is responsible for such a behavior of the magnetic field (by analogy with the Lenz law), additionally complicating the pinch effect.

6. CONCLUSIONS

Thus, we have obtained the following results. It is found that even and odd harmonics are generated during the motion of an electron in the skin layer produced by a relativistic laser pulse, the efficiency of generation of even harmonics being lower than the corresponding efficiency for odd harmonics. All harmonics are generated at the leading front of a laser pulse, the intensity of generation decreasing with increasing intensity of the laser pulse since electrons leave the skin layer. The generation efficiency decreases with increasing harmonic number. It should be noted that the harmonic generation efficiency also depends on the multiplicity of ionization of atomic ions.

The dynamics of electrons in the skin layer can also be analyzed using the equations of motion averaged over a laser period [17]. On the right-hand sides of these equations, ponderomotive forces appear on account of the coordinate dependence of the Hamilton function. Here, we are speaking of the dependence of the relativistic kinetic energy of an electron on the coordinate x along the direction of propagation of the laser beam in terms of the electric field amplitude $F^{in}(x)$, which decreases in the skin layer with increasing coordinate x. This dependence is the physical reason for the drift along the direction of propagation of a laser beam in the skin layer, which was considered above (see also [18]). However, harmonics should naturally be described using the equations of motion not averaged over a laser period.

In our calculations, we disregarded the decrease in the thickness of the skin layer due to relativistic effects (and the corresponding increase in the plasma frequency) [17]. The relativistic γ factor reduces the plasma frequency (see the Introduction), while the relativistic ponderomotive force increases this frequency. As a result, the plasma frequency increases as compared to the nonrelativistic value [17]. However, this increase is determined by electric field F^{in} in the skin layer rather than by electric field F in vacuum. When the condition $F^{in}/\omega c < 1$ is satisfied, this effect can be neglected, although $F/\omega c \gg 1$ for relativistic laser pulses.

ACKNOWLEDGMENTS

This study was supported financially by the Russian Foundation for Basic Research (project nos. 02-02-16678 and 04-02-16499) and the Basic Research and Higher Education Program (project no. MO-011-0).

REFERENCES

1. S. Banerjee, A. R. Valenzuela, R. C. Shah, *et al.*, Phys. Plasmas **9**, 2393 (2002).
2. V. P. Krainov, Phys. Rev. E **68**, 027401 (2003).
3. V. P. Kraĭnov and V. S. Rastunkov, Zh. Éksp. Teor. Fiz. **125**, 576 (2004) [JETP **98**, 508 (2004)].
4. S. V. Bulanov, T. Zh. Esirkepov, N. M. Naumova, and I. V. Sokolov, Phys. Rev. E **67**, 016405 (2003).
5. R. Ondarza-Rovira and T. J. M. Boyd, Phys. Plasmas **7**, 1520 (2000).
6. V. S. Rastunkov and V. P. Krainov, Phys. Rev. E **69**, 037402 (2004).
7. K. Krushelnick, I. Watts, M. Tatarakis, *et al.*, Plasma Phys. Controlled Fusion **44**, B233 (2002).
8. H. Hora, *Physics of Laser-Driven Plasmas* (Wiley, New York, 1981; Énergoatomizdat, Moscow, 1986).
9. V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics*, 4th ed. (Fizmatlit, Moscow, 2001; Pergamon Press, Oxford, 1982).
10. V. P. Silin, Kvantovaya Élektron. (Moscow) **27**, 283 (1999).
11. V. P. Silin, Zh. Éksp. Teor. Fiz. **114**, 864 (1998) [JETP **87**, 468 (1998)].
12. V. P. Silin, Zh. Éksp. Teor. Fiz. **117**, 926 (2000) [JETP **90**, 805 (2000)].
13. V. P. Silin, Kratk. Soobshch. Fiz. FIAN (Physical Inst., Russian Academy of Sciences) **8**, 32 (1998).
14. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 8th ed. (Nauka, Moscow, 2001; Pergamon Press, Oxford, 1975).
15. E. M. Lifshitz and L. P. Pitaevskiĭ, *Physical Kinetics*, 2nd ed. (Fizmatlit, Moscow, 2001; Pergamon Press, Oxford, 1981).
16. F. N. Beg, E. L. Clark, M. S. Wei, *et al.*, Phys. Rev. Lett. **92**, 095001 (2004).
17. R. N. Sudan, Phys. Rev. Lett. **70**, 3075 (1993).
18. N. E. Andreev, M. E. Veysman, V. P. Efremov, and V. E. Fortov, High Temp. **41**, 679 (2003).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 99, No. 3, 2004, pp. 494–503.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 3, 2004, pp. 566–576.

PLASMA, GASES

Distributions of Ions in a Cluster Plasma Created by a Laser Pulse

A. A. Katasonov[a] and M. B. Smirnov[a,b]

[a]*Moscow Institute of Physics and Technology, Dolgoprudnyĭ, Moscow oblast, 141700 Russia*
[b]*Russian Research Centre "Kurchatov Institute," Moscow, 123182 Russia*
e-mail: smirnov@imp.kiae.ru

Received March 19, 2004

Abstract—Energy and charge distributions of ions are calculated for a cluster beam irradiated by a high-power ultrashort laser pulse. It is shown that the self-consistent field of a cluster ionized by the laser beam strongly affects the characteristics of the ion distributions obtained after the cluster explodes. The mean concentration of atoms bound into clusters in a beam, the cluster size distribution, and the focal-spot diameter are found to have a weak effect on both energy and charge distributions of the ions, whereas the energy spectrum of the produced ions is determined by the mean cluster size. © *2004 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

The plasma created by irradiating a cluster beam with a high-power ultrashort laser pulse with a field strength higher than that characteristic of an atom (with pulse intensity $I \sim 10^{16}$–10^{20} W/cm^2 and pulse duration $\tau \sim 30$–300 fs) is a physical system whose evolution is determined by its inhomogeneity [1–3]. Under typical experimental conditions, the irradiated beam consists of clusters containing several thousand to several million atoms and is characterized by an average concentration of atoms varying between 10^{16} and 10^{19} cm^{-3}, which corresponds to a typical gas target. However, since the concentration of atoms in a cluster is high ($\sim 10^{22}$ cm^{-3}) [4], the density of the plasma of multicharged ions and electrons with energies on the order of several keV created inside the cluster substantially exceeds the cut-off electron density [5–7]. On the other hand, the electromagnetic wave freely propagates in this inhomogeneous medium and interacts with all target atoms, because the average electron density in the medium is lower than the cut-off density and the size of individual clusters is much smaller than the wavelength. Therefore, the laser pulse freely penetrates the regions where the electron density is higher than the cut-off density, whereas it is reflected from the dense homogeneous materials of solid-state targets [8]. Thus, a cluster beam can be used to reach high absorption efficiency, i.e., more efficient utilization of the target as compared to targets of other types.

For this reason, a cluster beam is advantageous as compared to gas and solid targets. First, the interaction of a laser pulse with a cluster beam results in almost complete (up to 95%) absorption of the pulse [9]. Second, the average ion charge exceeds the ion charge observed when either solid-state target or atomic beam is irradiated with a laser pulse characterized by similar parameters [10, 11]. To obtain an average ion charge comparable to that obtained by irradiating a cluster beam, the intensity of the laser pulse incident on a gas or solid target must increase by about an order of magnitude. Moreover, the pulse-energy fraction converted into x-rays can be as high as 10–15% and has a nonthermal nature [12–14]. The use of a deuterium molecular cluster beam as a target for the laser pulse provides a key neutron source where fusion involving deuterium ions takes place [15].

In this study, we calculate the charge and energy distributions for the multicharged ions created in the course of the cluster plasma decay resulting from the interaction between a cluster-ion beam and a laser pulse. This problem was computed in [16, 17] for an isolated cluster. In contrast, the present analysis takes into account the self-consistent field of a cluster ionized by a laser pulse, which is important for the process in question. Furthermore, with a view to simulating real experimental conditions, we allow for some additional factors that affect the process, including cluster size distribution and wave attenuation in the course of laser-pulse propagation across the cluster beam. A more realistic treatment of the problem makes it possible to achieve agreement between theoretical and experimental spectra of the ions produced.

2. MECHANISMS OF IONIZATION CAUSED BY INTERACTION BETWEEN A LASER PULSE AND A CLUSTER BEAM

Let us consider ionization and related processes caused by interaction between a laser pulse and a cluster beam. The ions created in solid-state and gas targets are accelerated by gradient forces, and their energy increases [18]. The mechanism of heating of the ions

created in a cluster beam is different. Each cluster in the beam interacts independently with the laser pulse, which leads to ionization of atoms in the cluster and ionization of the cluster as a whole. The resulting cluster has a positive charge and a sharp boundary. The cluster charge, as well as the charges of ions contained in the cluster, depends on the laser intensity, which decreases as the electromagnetic wave propagates across the cluster beam. Accordingly, the multicharged ions created by the laser pulse at the cluster boundary have the highest potential energy. The self-consistent electric field of the cluster accelerates the ions inside the cluster, which leads to its explosion. This leads to conversion of the potential energy of ions into kinetic energy, and the distribution of the resulting ions is almost isotropic in space and is an exponentially decreasing function of ion energy [19–21].

Ionization of atoms in a cluster interacting with a laser pulse leads to production of multicharged ions via two mechanisms: electron scattering by ions and direct ionization of atoms and ions by the electric field of the pulse. Depending on the cluster-beam and laser-pulse parameters, one of these mechanisms plays a dominant role. Ionization by electron impact dominates when the laser intensity is relatively low, not higher than 10^{16} W/cm^2 (i.e., when the electric-field strength is lower than that characteristic of an atom). As the charge of an ion increases, the contribution of this mechanism decreases. When the laser-pulse intensity is high and the ion charge reaches 8 to 10, the rate of electron impact ionization of atoms and ions virtually reduces to zero. In what follows, we focus on high-intensity laser pulses and ignore this mechanism of ionization of atoms and ions.

When the laser-pulse intensity is high, ions are produced as a result of ionization of cluster atoms and ions by the electric field of the electromagnetic wave, mainly via barrier-suppression ionization, because the tunneling of valence electrons takes a long time and can therefore be neglected. The electric-field strength G at which the potential barrier in the effective electron potential created by the Coulomb field of the core and the external electric field vanishes characterizes the binding energy of the valence electrons that can be released in this process. This field strength is given by Bethe's rule [22]:

$$G = \frac{J_Z^2}{kZ}, \tag{1}$$

where J_Z is the ionization potential for an ion with charge $Z - 1$ and k is a dimensionless factor, which is set equal to 4 as in [22]. Even though Bethe's rule is valid for quasi-electrostatic field, it holds for a varying field with a strength amplitude F if ionization can be considered instant, i.e., when the time required to release an electron is much shorter than the laser wave period ω^{-1}. This is true for barrier-suppression ionization if the Keldysh parameter γ is small,

$$\gamma = \frac{\omega\sqrt{2J_Z}}{F} \ll 1. \tag{2}$$

Using Bethe's rule (1), we rewrite (2) as follows:

$$\gamma \le \frac{2\omega Z^{1/4}}{F^{3/4}} \ll 1. \tag{3}$$

According to (3), the assumption of instant ionization is more reliable for stronger electric fields. For example, if the ion charge is $Z = 20$ atomic units (au), the electric-field strength is 5 au, and the electromagnetic-field frequency is 0.05 au, then the value of γ does not exceed 0.1. Under typical conditions of ionization of atoms and ions in a cluster, valence electrons are driven by both pulse electromagnetic field and self-consistent electric field of the charged cluster, whose strength is comparable to that of the laser field [23]. As a result, a stronger field acts on atoms and ions inside the cluster [11], and this effect is taken into consideration below.

In a cluster beam irradiated by a high-power laser, both individual ions and whole clusters are ionized concurrently. The mechanism of cluster ionization has much in common with ionization by electric field. Some of the electrons confined in the potential well created by the cluster field can break free by passing over the potential barrier when it is lowered by an external electromagnetic field. Treating the cluster as an ion, we can apply Bethe's rule (1) for ionization of an ion in electric field. Then, the cluster charge Q can be expressed in terms of F and cluster ionization potential J_Q as

$$Q = \frac{J_Q^2}{kF}. \tag{4}$$

Assuming that the ionization potential for a cluster of radius R is $J_Q = Q/R$, we obtain

$$Q = kFR^2. \tag{5}$$

As in the case of an isolated atom, Bethe's rule holds for cluster ionization if it can be considered as instant. Since the electron velocity can be estimated as $\sqrt{Q/R}$, and the time required for an electron to travel a distance comparable to R is

$$t_{\text{dis}} \sim \sqrt{\frac{R^3}{Q}}, \tag{6}$$

the condition $t \ll 1/\omega$ for the field to be treated as quasi-

electrostatic becomes

$$R^3\omega^2 \ll Q. \quad (7)$$

Using expression (5) for the cluster charge, we rewrite criterion (7) as

$$\frac{R\omega^2}{4F} \ll 1. \quad (8)$$

In particular, criterion (7) is satisfied for a laser field strength of 5 au, a pulse wave frequency of 0.05 au, and a cluster radius of 500 au:

$$\frac{4F}{R\omega^2} = 80 \gg 1.$$

Therefore, expression (5) for the charge of an ionized cluster holds in a wide range of cluster-beam and laser-pulse parameters.

Thus, irradiation of a cluster with a high-power ultrashort laser pulse results in its excitation and partial ionization. An excited cluster consists of multicharged ions and electrons confined in the self-consistent cluster field. This unstable system explodes in several tens of picoseconds. The explosion of a cluster is caused either by the action of the self-consistent cluster field on the cluster ions or by the pressure exerted by the electron gas on the ions. The latter scenario is possible only under special conditions [23] and is neglected here. In the former explosion scenario, the potential energy of an ion with charge Z in the self-consistent cluster field,

$$E = Z\frac{Qr^2}{R^3}, \quad (9)$$

is completely converted into kinetic energy after interaction with a laser pulse. The resulting kinetic energy of an ion depends on its charge, distance r from the center of the cluster, and cluster size R. Expression (9) holds when ions of mass M remain at rest during the interaction with the laser pulse, i.e., if the displacements of ions on the cluster surface are small as compared to the cluster size. This is true for pulses having durations such as

$$\omega\tau \ll \sqrt{\frac{R\omega^2}{F}}\sqrt{\frac{Z}{M}}. \quad (10)$$

In particular, for a laser pulse of intensity 10^{18} to 10^{19} W/cm^2 interacting with a xenon beam, the criterion is satisfied if the pulse duration does not exceed 10 to 20 laser cycle periods (30 to 100 fs).

3. CALCULATION OF ION DISTRIBUTIONS

To find the energy and charge distributions of ions produced when a high-power ultrashort laser pulse is absorbed by a cluster beam, we analyze the first stage of the process, when the electromagnetic interacts with each cluster independently. Since the density of the plasma obtained as a result of cluster explosions is sufficiently low for ion–ion collisions to be negligible on a relatively large time scale, we assume that the ion distribution after the explosions of individual clusters is similar to that observed in the resulting homogeneous plasma. At the initial stage of ion production under these conditions, bound electrons are released via barrier-suppression ionization caused by the combined effect of the pulse field and the self-consistent cluster field. Some electrons escape to infinity, and the cluster acquires a positive charge. The remainder of the released electrons is confined in the self-consistent cluster field. Both the cluster charge and the charges of ions in the cluster after the interaction with the laser pulse determine the energy and charge distributions of the ions after the cluster explosions, depending on the cluster size and the laser-pulse intensity. The model proposed here to describe the interaction between a laser pulse and a cluster beam makes use of both cluster size distribution and variation of laser intensity in space and time.

It is important for the present analysis that the cluster beam contains clusters of different size. Relying on experimental results, we use here both lognormal and exponential cluster size distributions in a cluster beam [4],

$$f(R) = c_1\exp\left(-\frac{1}{\delta^2}\ln^2\frac{R}{\bar{R}}\right) \quad (11)$$

and

$$f(R) = c_2R^5\exp\left(-\frac{R^3}{\bar{R}^3}\right), \quad (12)$$

where $\bar{R}$ is the mean cluster size in the beam and δ denotes the width of a lognormal distribution.

The pulse field strength determining the cluster and ion charges depends both on the location of the cluster in the focal spot and on the time when the pulse begins to interact with a particular cluster. As the electric field penetrates to a distance l in a cluster beam, its strength exponentially decreases, with a decay rate constant β depending on the cluster density:

$$F = F_0\exp(-\beta l). \quad (13)$$

The laser-pulse intensity decreases with increasing distance from the laser axis. From here on, we use a Gaus-

sian field-strength distribution along the distance from the axis:

$$F \sim \exp\left[-\left(\frac{\rho}{\rho_0}\right)^2\right], \tag{14}$$

where ρ_0 is the radius of the focal spot. Combining these expressions, we find the field strength at a point $\mathbf{X} = \{l, \rho, \phi\}$ in the laser pulse (ϕ is an arbitrary azimuthal angle):

$$F = F_0 \exp\left[-\beta l - \left(\frac{\rho}{\rho_0}\right)^2\right] = F_0 \Phi(\mathbf{X}), \tag{15}$$

where F_0 is the highest field strength and the function $\Phi(\mathbf{X})$ varies from 0 to 1. Note also that, by definition, the function $\Phi(\mathbf{X})$ has an important property characterizing the spatial distribution of the field strength:

$$d\Phi = -\beta dl 2\pi\rho d\rho \frac{1}{\rho_0^2}\Phi\{\mathbf{X}\} = -\alpha\Phi\{\mathbf{X}\}d\mathbf{X}, \qquad \alpha = \frac{2\pi\beta}{\rho_0^2}, \tag{16}$$

where $d\mathbf{X}$ denotes a volume element in the cluster beam.

Let us find the energy and charge distributions for ions produced after the cluster explodes at the first stage of the formation of a homogeneous plasma. Ion–ion collisions in the resulting plasma are neglected because its density is relatively low (the mean free paths of ions are large). The charges of the ions created in the cluster are also uniquely determined, because the ionization potential for electrons to be released by barrier suppression is related to the field strength in a laser pulse propagating in the region where the cluster is located by Bethe's rule. The energy of an ion located at a certain distance r from the cluster's center is also determined by (9) as a function of r and Q if the charge distribution in the cluster is uniform. Therefore, both energy and charge distributions of the ions produced in a cluster explosion are determined by the intensity of the electromagnetic wave incident on a cluster and the cluster radius.

Thus, the determination of the energy and charge distributions for ions produced after the cluster explosions caused by interaction with a laser pulse is a readily solvable, but tedious task, because both charge and energy of an ion depend on five variables: the three coordinates of the exploding cluster, the cluster size, and the distance from the ion to the cluster's center. The desired result must be averaged over these five variables by using the distributions specified above. The averaging procedure makes use of a simplified average-ion model in which both the highest ionization potential for electrons released via barrier suppression and the charge of the resulting ions are continuous functions of laser-pulse intensity [24, 25].

To perform the averaging, we represent the number of ions in a cluster of radius $R \in [R, R + dR]$ located at a distance $r \in [r, r + dr]$ from the center of the cluster in a volume element $\mathbf{X} \in [\mathbf{X}, \mathbf{X} + d\mathbf{X}]$ of the cluster beam as

$$dN = n_i 4\pi r^2 dr N_{\mathrm{cl}} f(R) dR d\mathbf{X}. \tag{17}$$

Here, n_i is the concentration of ions in a cluster, N_{cl} is the concentration of clusters in the beam (assumed to be spatially uniform within the beam), and $f(R)$ is the cluster size distribution. Actually, the problem reduces to representation of dN in (17) as a function of ion energy and charge. We use (1), (9), and (15) to express r and Φ in terms of the charge and potential energy of an ion:

$$r = \left(\frac{4ER}{kJ_Z^2}\right)^{1/2}, \quad \Phi = \frac{J_Z^2}{kZF_0}. \tag{18}$$

In the new variables, the required ion concentration is written as

$$dN = 4\pi n_i \frac{4ER}{kJ_Z^2} N_{cl} f(R)\alpha dR \frac{d\Phi(Z, E)}{\Phi(Z, E)} dr(Z, E). \tag{19}$$

The differentials dr and $d\Phi$ are expressed in terms of dZ and dE to obtain

$$\frac{d^2N}{dEdZ} = \begin{cases} A\dfrac{\sqrt{E}}{J_Z^3}\left(\dfrac{2J_Z'}{J_Z} - \dfrac{1}{Z}\right)\displaystyle\int_{R_{\min}}^{\infty} R^{3/2} f(R) dR, & Z < Z_{\max}, \\ 0, & Z > Z_{\max}, \end{cases} \tag{20}$$

where

$$A = \frac{16\pi n_i N_{\mathrm{cl}}\alpha}{k^{3/2}} \tag{21}$$

is a constant that determines the absolute ion yield and $Z_{\max}$ is the largest ion charge corresponding to a particular intensity, which can be found by using Bethe's rule (1).

Note that the integral with respect to cluster size in (20) is performed over the interval

$$R \in \left[R_{\min} = \frac{4E}{kJ_Z^2}, \infty \right].$$

Even though this follows from mathematical calculations, the contribution of small clusters to the distribution of ions is limited for physical reasons. Indeed, using (9) and (4), we find that an energy larger than $kJ_Z^2 R_{\min}/4$ cannot be transferred from a cluster of radius smaller than $R_{\min}$ to an ion. Calculating the integral in (20), we obtain expressions corresponding to the log-normal and exponential cluster size distributions, respectively:

$$\frac{d^2N}{dEdZ} \sim \begin{cases} \frac{\sqrt{E}}{J_Z^3}\left(\frac{dJ_Z}{dZ}\frac{2}{J_Z} - \frac{1}{Z}\right)\left[1 - \mathrm{erf}\left(\ln\left(\frac{4E}{k\delta\bar{R}J_Z^2} - \frac{5\delta}{4}\right)\right)\right], & Z < Z_{\max}, \\ 0, & Z > Z_{\max}, \end{cases} \tag{22}$$

$$\frac{d^2N}{dEdZ} \sim \begin{cases} \frac{\sqrt{E}}{J_Z^3}\left(\frac{dJ_Z}{dZ}\frac{2}{J_Z} - \frac{1}{Z}\right)\left[e^{-\gamma}\sqrt{\frac{\gamma}{\pi}}\left(2 + \frac{\gamma}{3}\right) + 1 - \mathrm{erf}(\sqrt{\gamma})\right], & Z < Z_{\max}, \\ 0, & Z > Z_{\max}, \end{cases} \tag{23}$$

where

$$\gamma = \left(\frac{4E}{kJ_Z^2\bar{R}}\right)^3.$$

The self-consistent field of a charged cluster increases the ion charge, modifying the ion distribution. Taking into account the contribution of the self-consistent cluster field to the total strength of the field that ensures barrier-suppression ionization of atoms in the cluster, we rewrite Bethe's rule as

$$F + \frac{Qr}{R^3} = F\left(1 + \frac{kr}{R}\right) = \frac{J_Z^2}{4Z}. \tag{24}$$

Again, the field strength F in the electromagnetic wave depends on the location of cluster in the focal spot of the laser pulse according to (16). To simplify calculations, we change from R and r to the dimensionless coordinates x and y normalized to the mean cluster radius in the beam,

$$x = \frac{r}{\bar{R}}, \quad y = \frac{R}{\bar{R}}. \tag{25}$$

Then, expression (9) for energy and Bethe's rule (24) become

$$\frac{E}{kZF_0\Phi\bar{R}} = \frac{x^2}{y}, \tag{26}$$

$$\left(\frac{J_Z^2}{4ZF_0\Phi} - 1\right)\frac{1}{k} = \frac{x}{y}. \tag{27}$$

Using the parametrization

$$x = \frac{kC+1}{C}\beta, \quad y = \frac{kC+1}{C^2}\beta, \quad \beta = \frac{4E}{k\bar{R}J_Z^2} \tag{28}$$

in (26) and (27), we obtain

$$\frac{E}{kZF_0\Phi\bar{R}} = \beta(kC+1), \quad C = \left(\frac{J_Z^2}{4ZF_0\Phi} - 1\right)\frac{1}{k}. \tag{29}$$

According to (17), the number of ions with energy E and charge Z is

$$\frac{d^2N}{dZdE} \sim \begin{cases} \frac{4\pi n_i N_{\mathrm{cl}}\alpha\bar{R}^4}{k}\int_0^1 f(\bar{R}y(C))\frac{\beta^4}{E}\left(\frac{2J_Z'}{J_Z} - \frac{1}{Z}\right)\frac{(kC+1)^5}{C^6}\frac{d\Phi}{\Phi}, & Z < Z_{\max}, \\ 0, & Z > Z_{\max}, \end{cases} \tag{30}$$

Using the definition of C and Eq. (24), we obtain

$$\frac{kdC}{kC+1} = -\frac{d\Phi}{\Phi}, \tag{31}$$

which makes it possible to rewrite (30) as

$$\frac{d^2N}{dEdZ} \sim \begin{cases} 4\pi n_i N_{\text{cl}} \alpha \bar{R}^4 \frac{\beta^4}{E}\left(\frac{dJ_Z}{dZ}\frac{2}{J_Z} - \frac{1}{Z}\right)\int\limits_0^1 f(R(C))\frac{(kC+1)^4}{C^6}dC, & Z < Z_{\max}, \\ 0, & Z > Z_{\max}. \end{cases} \tag{32}$$

By virtue of (31) and the properties of the function Φ introduced in (15), it holds that $C \in [0, 1]$. Changing from C to variable y related to cluster size and discarding a numerical factor independent of ion energy and charge, we obtain

$$\frac{d^2N}{dEdZ} \sim \begin{cases} \frac{1}{Ea}\left(\frac{dJ_Z}{dZ}\frac{2}{J_Z} - \frac{1}{Z}\right)\int\limits_{a\frac{k+1}{k^2}}^{\infty} \frac{(a+(a+4ya)^{1/2})^5}{a+4y+(a^2+4ya)^{1/2}}dyf(y), & Z < Z_{\max}, \\ 0, & Z > Z_{\max}, \end{cases} \tag{33}$$

$$a = \beta k^2 = \frac{4kE}{\bar{R}J_Z^2}.$$

Note that $Z_{\max}$ is given by Eq. (24), where the self-consistent cluster field is taken into account. Thus, real distributions of laser-pulse intensity and relations for the charge and energy of an ion can be used as analytical expressions to reduce the number of parameters in the averaging procedure from five to one. This simplifies mathematical treatment, while a more or less realistic model is retained.

4. NUMERICAL RESULTS AND DISCUSSION

Let us analyze the numerical results obtained by implementing the program presented above. Note that the highest ion energies correspond to the focal spot and the ions created on the cluster surface. Figure 1 shows the largest ion charge calculated as a function of laser-pulse intensity for a cluster radius of 500 au by taking into account the self-consistent cluster field. In what follows, we discuss ion distributions computed for typical experimental values of parameters: for a xenon cluster beam with $\bar{R}$ = 500 au, δ = 0.2 (see (11)), and k = 4 and a laser-pulse intensity of 8×10^{17} W/cm^2 (F = 5 au). These parameters correspond to the largest ion charge Z = 18 when the self-consistent field in a charged cluster is neglected and Z = 26 under real conditions when the field is taken into account. The highest ion energy corresponding to a cluster of radius R = 26.5 nm and the laser-pulse intensity indicated above is several MeV.

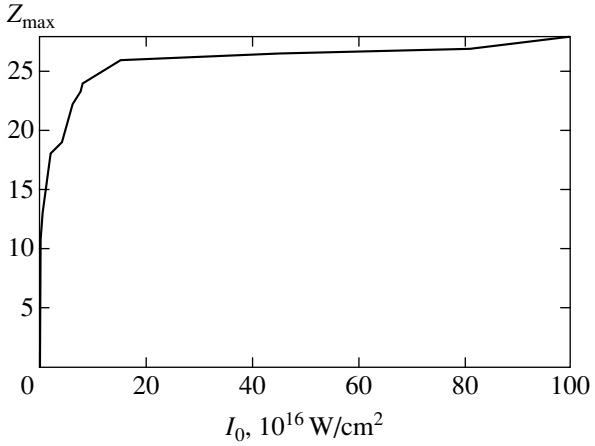

Fig. 1. Largest xenon ion charge versus laser-pulse intensity.

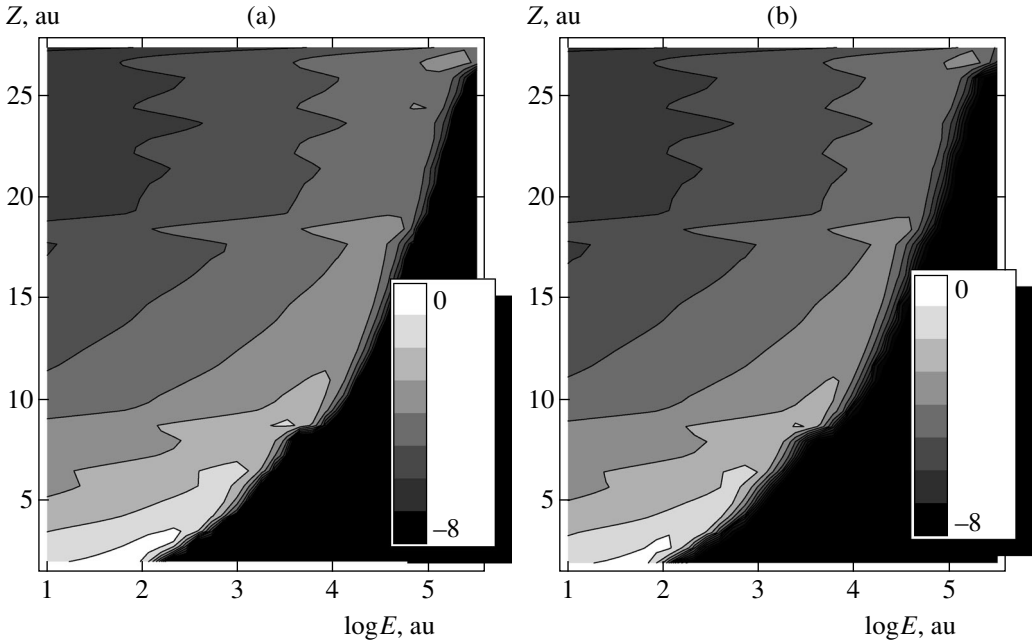

Fig. 2. Xenon ion charge–energy distributions corresponding to lognormal (a) and exponential (b) cluster size distributions for $\bar{R}$ = 26.5 nm and a laser intensity of 8×10^{17} W/cm^2.

Figure 2 shows the ion distributions obtained under the conditions specified above for lognormal and exponential cluster size distributions. The areas corresponding to charges and energies characterized by probabilities differing by an order of magnitude are separated by curves. Lighter areas correspond to higher values of ion distribution functions. Comparing the ion distributions, one finds that they are qualitatively similar: both distributions of ions reach maxima and minima at the same values of charge and energy. The distributions are identical almost in the entire domain. However, the ion and energy distributions have sharper boundaries for the exponential cluster size distribution as compared to the lognormal one. This is demonstrated by Fig. 3 for the ion energy and charge distribution function corresponding to Z = 12. Note that allowance for the self-consistent cluster field during ionization is essential for both charge and energy distributions of ions. In particular, when the laser-pulse intensity is 8×10^{17} W/cm^2 as indicated above, the largest ion charge is 18 instead of 26 if the self-consistent field is neglected, while the highest ion energies calculated with and without allowance for the self-consistent field differ by a factor of 1.3. Note that the largest charge of the created ions is independent of the cluster size and is determined solely by the laser-pulse intensity [11].

Figure 4 shows an ion distribution function represented by a three-dimensional graph. The function is localized in a narrow charge–energy area with sharp boundaries. For a constant charge, this corresponds to an exponentially decreasing ion distribution function at low energies and nearly zero values of the function at high energies. With increasing ion charge, the peak value decreases everywhere except for Z = 8, 18, and 26, which correspond to closed-shell ions. In these special cases, the ionization potential has sharp peaks as a function of ion charge, and production of ions with large charges requires a high-intensity laser pulse. This behavior of distributions was observed experimentally in [10], but the peaks obtained in the present study are sharper. Thus, there exists an area in the plane of Z and

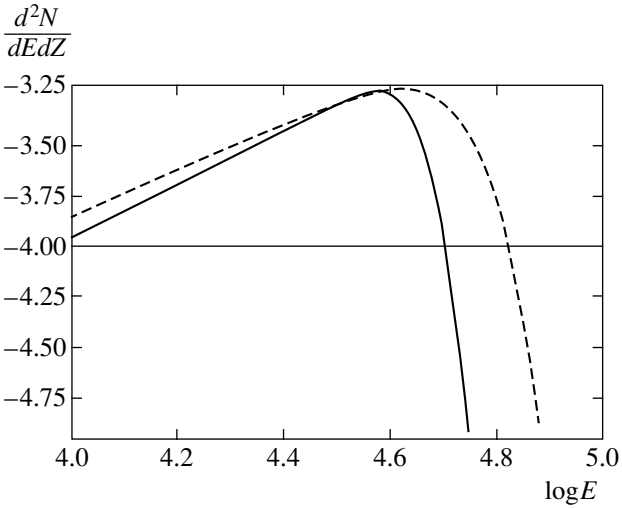

Fig. 3. Xenon ion energy distributions for $\bar{R}$ = 26.5 nm and a laser intensity of 8×10^{17} W/cm^2 at Z = 12. Solid and dashed curves correspond to lognormal and exponential cluster size distributions, respectively.

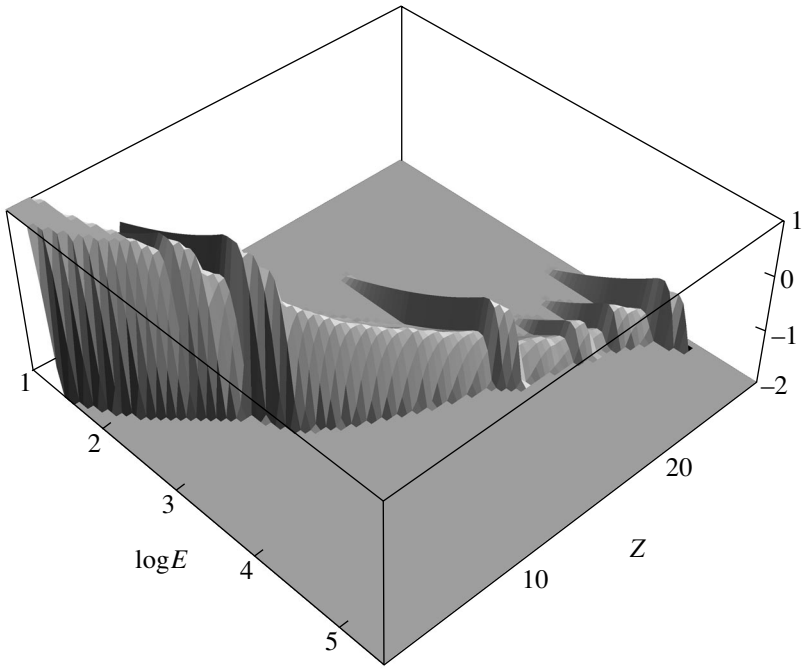

Fig. 4. Ion charge–energy distribution corresponding to a lognormal xenon-cluster size distribution for $\bar{R}$ = 26.5 nm and a laser intensity of 8×10^{17} W/cm^2.

E where the ion distribution function substantially differs from zero.

Let us consider the area where the ion charge–energy distribution function rapidly decreases with increasing energy. In Fig. 2, this area is characterized by the highest density of isopleths. It is obvious that there exist extreme values of energy and charge corresponding to a specific cluster size. However, if the cluster size distribution includes clusters of different size, then these extreme values should not be well defined. Nevertheless, computations reveal a sharp boundary of the distribution function at high values of energy and charge, as illustrated by Fig. 2. Note that this boundary is independent of the laser-pulse intensity that corresponds to the largest possible charge. The boundary is given by the relation

$$\bar{R}C'J_Z^2 = \log E. \tag{34}$$

This behavior is explained by the exponential form of the cluster size distribution in the large-size limit. According to our calculations, the numerical factor in (34) is $C' = 9/20$.

Figure 5 shows the ion energy spectra obtained by integrating the charge–energy distributions with respect to charge. It demonstrates that the spectra corresponding to lognormal and exponential cluster size distributions are similar. Furthermore, these spectra agree with those found experimentally in [19–21] in the interval of ion energy between ~1 keV and ~1 MeV. The results predicted by the present theory for ion energies below 1 keV differ from experiment, because we ignore collisions.

Note that the ion energy spectrum corresponds to the peak values of the charge–energy distribution function. This is illustrated by Fig. 6, where dashed curves represent the ion energy spectra corresponding to the specified values of cluster-beam and laser-pulse param-

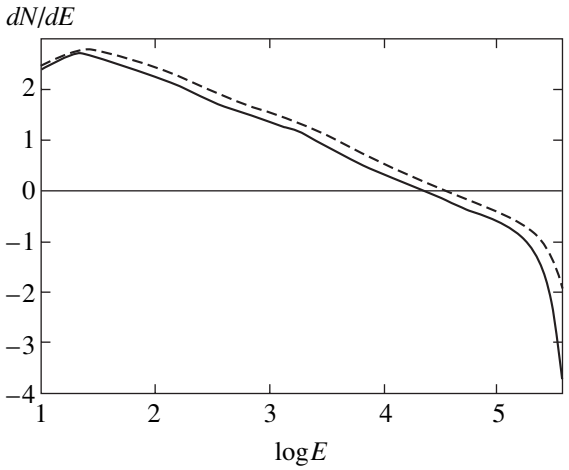

Fig. 5. Xenon ion energy spectra for lognormal (solid curve) and exponential (dashed curve) cluster size distributions for $\bar{R}$ = 26.5 nm and a laser intensity of 8×10^{17} W/cm^2.

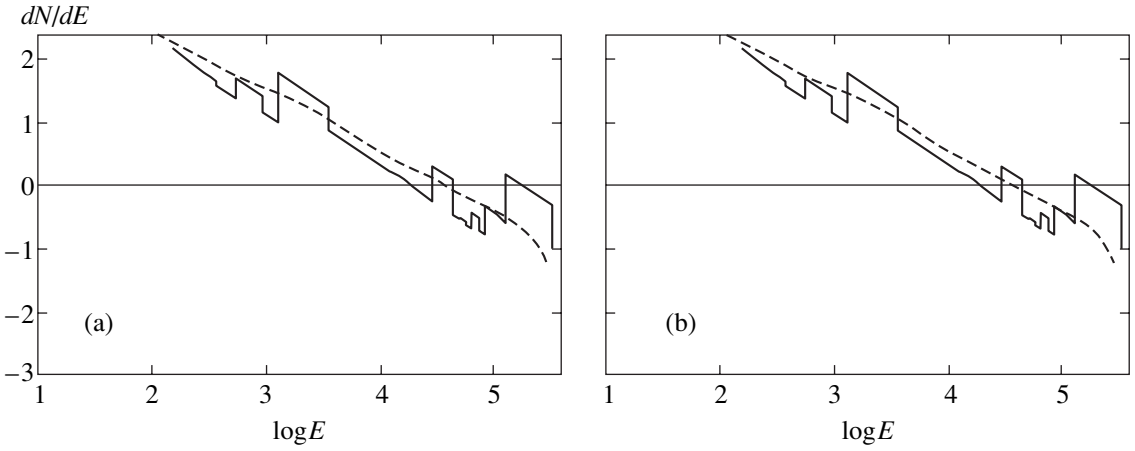

Fig. 6. Ion energy spectra for $\bar{R}$ = 26.5 nm and a laser intensity of 8×10^{17} W/cm^2 (dashed curve) and peak of charge–energy distribution function multiplied by the characteristic width (solid curve) corresponding to lognormal (a) and exponential (b) cluster size distributions.

eters. It is clear that the ion energy spectrum is close to the peak values of the charge–energy distribution function if this function is multiplied by the characteristic width of the ion charge distribution function.

5. CONCLUSIONS

Our computations of energy and charge–energy distributions for ions created by irradiating a cluster beam with a high-power ultrashort laser pulse demonstrate the important role played by the self-consistent field of an exploding cluster in the formation of the ion spectrum. According to the computations, the distributions are similar in shape, and their characteristics are primarily determined by the mean cluster size and the largest ion charge, which depend on the laser-pulse intensity in the focal spot at the cluster-beam boundary, where the laser-pulse intensity is the highest. The ion distributions computed for exponential and lognormal cluster size distributions differ only for charge and energy values characterized by low probabilities. The computations demonstrate that the dominant components of the resulting plasma are closed-shell ions. When the self-consistent cluster field is taken into account, the shapes of the ion distributions do not change, but the ionization efficiency increases, and so does the largest charge. It can be expected that any model ignoring the effect of the self-consistent field will lead to qualitatively correct results that are quantitatively inconsistent with experiment.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 04-02-16499 and CRDF, grant no. MO-011-0.

REFERENCES

1. T. Ditmire, Contemp. Phys. **38** (5), 315 (1998).
2. B. D. Thompson, A. McPherson, K. Boyer, and C. K. Rhodes, J. Phys. B **27**, 4391 (1994).
3. Ş. Dobosz, M. Schmidt, M. Perdrix, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **68**, 454 (1998) [JETP Lett. **68**, 485 (1998)].
4. B. M. Smirnov, *Clusters and Small Particles in Gases and Plasmas* (Springer, New York, 2000).
5. T. Ditmire, T. Donnelly, A. M. Rubenchik, *et al.*, Phys. Rev. A **53**, 3379 (1996).
6. V. P. Kraĭnov and M. B. Smirnov, Usp. Fiz. Nauk **170**, 969 (2000) [Phys. Usp. **43**, 901 (2000)].
7. V. P. Krainov and M. B. Smirnov, Phys. Rep. **370**, 237 (2002).
8. H. M. Michelberg and R. R. Freeman, J. Opt. Soc. Am. B **6**, 1351 (1989).
9. Y. L. Shao, T. Ditmire, J. W. G. Tisch, *et al.*, Phys. Rev. Lett. **77**, 3343 (1996).
10. M. Lezius, S. Dobosz, D. Normand, and M. Schmidt, Phys. Rev. Lett. **80**, 261 (1998).
11. M. B. Smirnov and V. P. Krainov, Phys. Rev. A **69**, 043201 (2004).
12. S. Ter-Avetisyan, M. Schnürer, H. Stiel, *et al.*, Phys. Rev. E **64**, 036404 (2001).
13. M. Mori, T. Shiraishi, E. Takahashi, *et al.*, J. Appl. Phys. **90**, 3595 (2001).
14. T. Mocek, C. M. Kim, H. J. Shin, *et al.*, Appl. Phys. Lett. **76**, 1819 (2000).
15. J. Zweiback, T. E. Cowan, R. A. Smith, *et al.*, Phys. Rev. Lett. **85**, 3640 (2000).
16. I. Last and J. Jortner, Phys. Rev. A **62**, 013201 (2000).
17. M. Eloy, R. Azambuja, J. T. Mendonca, and R. Bingham, Phys. Plasmas **8** (3), 1084 (2001).

18. T. Zh. Esirkepov, S. Bulanov, H. Daido, *et al.*, Phys. Rev. Lett. **89**, 175003 (2002).
19. T. Ditmire, E. Springate, J. W. G. Tisch, *et al.*, Phys. Rev. A **57** (1), 369 (1998).
20. Lei An-Le, Li Zhong, Ni Guo-Quan, and Xu Zhi-Zhan, Chin. Phys. **9** (6), 432 (2000).
21. K. W. Madison, K. Patel, M. Allen, *et al.*, J. Opt. Soc. Am. B **20** (1), 113 (2003).
22. H. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms*, 2nd ed. (Rosetta, New York, 1977; Fizmatgiz, Moscow, 1960).
23. M. B. Smirnov, Zh. Éksp. Teor. Fiz. **124**, 48 (2003) [JETP **97**, 42 (2003)].
24. N. E. Andreev, I. L. Beĭgman, M. E. Veĭsman, *et al.*, Preprint No. 59, FIAN (Physical Inst., Russian Academy of Sciences, Moscow, 1998).
25. M. B. Smirnov, I. Yu. Skobelev, A. I. Magunov, *et al.*, Zh. Éksp. Teor. Fiz. **125**, 1283 (2004) [JETP **98**, 1123 (2004)].

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 3, 2004, pp. 504–509.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 3, 2004, pp. 577–584.

FLUIDS

Physical Processes Underlying the Development of Shear Turbulence

O. M. Belotserkovskiĭ[a], A. M. Oparin[a], and V. M. Chechetkin[b]

[a]*Institute for Computer-Aided Design, Russian Academy of Sciences, ul. Vtoraya Brestskaya 19/18, Moscow, 123056 Russia*

[b]*Institute of Applied Mathematics, Russian Academy of Sciences, Miusskaya pl. 4, Moscow, 125047 Russia*

e-mail: a.oparin@icad.org.ru

Received January 22, 2004

Abstract—A physical model of the development of turbulence in free shear flows is proposed. The model is based on the results of numerical simulations of turbulent flow development. The main ideas of the proposed theory of turbulence are stated as follows: the onset of turbulence begins with the formation of large vortices; spectral energy transfer involves both direct and inverse cascades; and the inertial range of the energy spectrum develops as a result of concurrent direct and inverse cascades. The dominant physical factors that determine the spectrum include Joukowski forces. © *2004 MAIK "Nauka/Interperiodica".*

1. BACKGROUND

The problem of turbulence has challenged scientists for over a century. However, no comprehensive turbulence theory has been developed to this day. Let us reiterate the basic ideas that have been considered.

Starting from the work of O. Reynolds, the theory of fully developed turbulence relies on the assumption that flow velocity is a Gaussian random variable fluctuating about a mean value at any point. This implies that the velocity length-scale (wavenumber) distribution is consistent with an appropriate mathematical expectation. This hypothesis constitutes the basis for the statistical theory of turbulent flows. In 1925, Keller and Friedmann introduced a general definition of Eulerian space-time correlation functions for turbulent flow fields and suggested a method for deriving time-dependent equations for correlation functions from the Navier–Stokes equations. These equations play a fundamental role in the statistical turbulence theory. The most comprehensive presentations of the theory can be found in [1, 2].

The equations for correlation functions are used to average flow characteristics, such as velocity or vortex size. The mathematical model assumes that flow perturbations are random and can be described by probabilistic methods extended to function spaces. In fact, this assumption underlies the derivations of simplified equations and estimates for the growth exponents of the instabilities leading to the onset of turbulence.

However, the velocity measured at a particular point in a turbulent flow mostly remains within the limits corresponding to large-scale vortices, because these vortices are the major elements of flow structure. Fast fluctuations can be observed when the small eddies surrounding large-scale vortices move past the velocity sensor.

A variety of subgrid-scale models have been developed for applied simulations of turbulent flows, including the k–ε model, Reynolds stress models, and large-eddy simulation (LES) models. All of these models contain semiempirical constants that can be adjusted to compute particular flows.

In 1985, O.M. Belotserkovskiĭ showed that fully-developed free turbulence could be simulated without using any subgrid-scale model and adjusting any semiempirical constants [3]. The starting point of this line of research was a von Kármán lecture published in [4]. The most systematic presentations of this approach can be found in [5, 6], where it was applied to free turbulent flows behind moving bodies (including both near- and far-wake flow structures), oceanic flows, Taylor–Couette flow, evolution of turbulent mixing zones, and other important problems concerning the onset of turbulence. The approach reflects the multidimensional and unsteady nature of the flows in question and takes into account phenomena related to compressibility, as well as effects due to viscosity (dominated by molecular mechanism). In those studies, it was also shown that large-scale vortices play a dominant role in turbulent flow structure. It should also be noted that this idea was supported by A.S. Monin [7].

Much later, Western specialists proposed the Monotonically Integrated Large Eddy Simulation approach (known as the MILES model) [8]. As in [3], the Euler or Navier–Stokes equations were solved by using upwind monotone finite-difference schemes without any subgrid-scale modeling. Monotone differencing introduces a dissipative mechanism (different from physical viscosity) that implicitly plays the role of a

subgrid-scale model, while the large-scale motion resolved on the grid is assumed to be independent of the actual mechanism of small-scale dissipation.

2. BASIC IDEAS

The understanding of the onset of turbulence was greatly improved by recognizing the key role played by large-scale coherent structures. In this context, the statistical treatment of turbulence based on a probabilistic description must be combined with deterministic modeling of evolution of large-scale vortices. Physical models of this kind can be implemented in direct numerical simulations [3], where the development of flow disturbances on progressively smaller scales is due to interactions of large-scale vortices with the background flow and with one another, rather than to bifurcations of trajectories in the phase space [9].

The basic ideas of direct numerical simulation of turbulence rely on the following two hypotheses supported by experimental evidence:

(1) large-scale coherent vortices and small-scale stochastic turbulence are statistically independent at high Reynolds numbers;

(2) molecular viscosity (more generally, the mechanism of energy dissipation) plays a minor role in the analysis of large-scale vortex dynamics.

Large vortices carry the greater part of the energy of turbulent motion and determine the flow structure. The dynamics of large vortices does not reflect the structure of random fluctuations, being governed by the Navier–Stokes equations in which the inertial terms dominate over the viscous terms. Accordingly, the structure of a vortex develops as a result of combined action of pressure gradients and transient forces arising from velocity fields. Therefore, both formation of large-scale vortices and ensuing flow structure must be described by the Euler equations.

Molecular viscosity is taken into account only in analyzing small-scale turbulent motion and dissipation of kinetic energy into heat. However, energy is dissipated only in small eddies with length scales close to the spectral boundary. Therefore, the distribution of thermal-energy production rate over wave numbers is not isotropic. Further evolution of a large-scale vortex leads to generation of small-scale eddies at the boundary between the vortex and the background flow, i.e., to intermittent behavior of turbulence. We conclude that local flow patterns in well-developed turbulence are unsteady and small-scale components of turbulent spectra must depend on the mechanism of vorticity generation in the neighborhoods of large-scale vortices.

To examine physical scenarios of the onset of turbulence, we performed an extensive series of numerical simulations of free shear flows of an inviscid perfect gas. We used monotone dissipative stable finite-difference schemes with positive operators [3–6]. This technique requires neither special filtering nor semiempirical modeling to approximate subgrid turbulence. (It should be noted here that the inertial-range spectrum was obtained by Kolmogorov without assuming any specific form of dissipative terms.) The schemes automatically introduce a dissipative mechanism playing the role of a subgrid turbulence model.

We analyzed the evolution of a shear layer with a uniform velocity gradient. Figure 1 demonstrates that large vortices of diameter comparable to the shear-layer thickness develop first. The vortex motion in a finite volume is generated by a pressure gradient. The shear layer breaks up into large vortices, and smaller eddies develop in their wakes. At the final instant of the simulation, the flow consists of a single vortex occupying the computational domain. This effect is explained by the attraction of vortices with similar vorticity signs due to the Joukowski force (see Fig. 2). The computation was performed with free-flow conditions set on the upper and lower boundaries combined with periodic conditions at the upstream and downstream boundaries.

Note that a single large-scale vortex (Great Red Spot) persists in the Jovian atmosphere, as illustrated by Fig. 3.

Figure 4 shows the results obtained by computing the evolution of turbulence in a similar shear layer, but with impermeability conditions set on the upper and lower boundaries. Here, the onset of turbulence follows the scenario observed in the preceding simulation. However, the final turbulent flow has a complicated pattern involving both large-scale vortices and smaller structural elements. This result is due to interactions of the background flow with walls and large-scale vortices.

In 1920, L.F. Richardson pointed to the key role played by an energy cascade in turbulence structure. Indeed, different turbulent energy spectra can be obtained, depending on the mechanism of the cascade. The classical Kolmogorov–Obukhov spectrum was obtained by Kolmogorov by assuming that the rate of spectral energy transfer is constant and the process of eddy breakup is self-similar, while viscous dissipation of kinetic energy into heat occurs only in small-scale eddies. Note that it has long since been understood that only small-scale components of turbulent motion are fully stochastic. Large vortices are coherent and cannot be described in terms of random fields. While it is clear that the conversion of kinetic energy into heat is due to viscosity, it is important to find out whether the energy of large vortices or the energy of vortex–flow interaction is responsible for the generation of smaller vortices. The flow energy may play an important role since large vortices are relatively stable.

Analysis of the processes contributing to the development of the turbulent energy spectrum is the pivotal element of turbulence theory. It is generally believed that the spectrum develops as a result of energy transfer from large-scale vortices (as in the Kolmogorov–Obukhov turbulence theory). However, it was shown in a recent analysis that an important role is played by

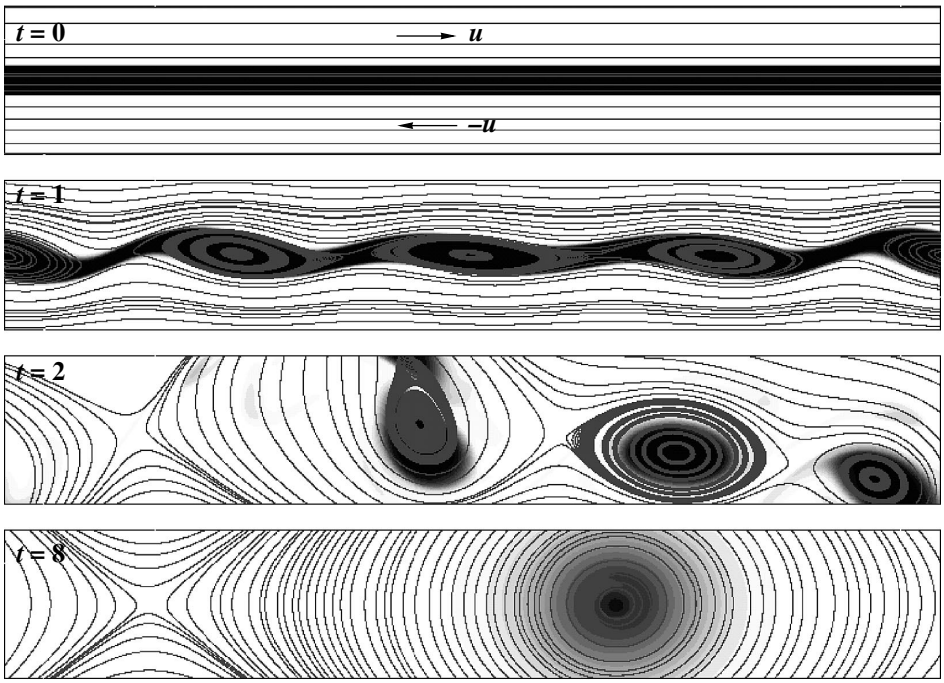

Fig. 1. Development of large-scale vortices in a free turbulent shear layer. Streamlines are shown at instants separated by equal time intervals, including the starting moment. The grayscale value represents the concentration of particles initially localized in the shear layer.

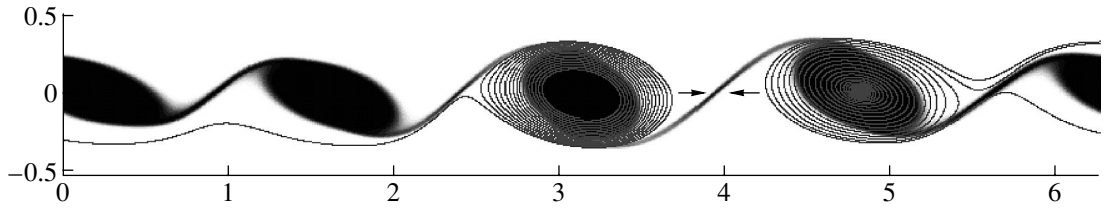

Fig. 2. Attraction of vortices with similar vorticity signs.

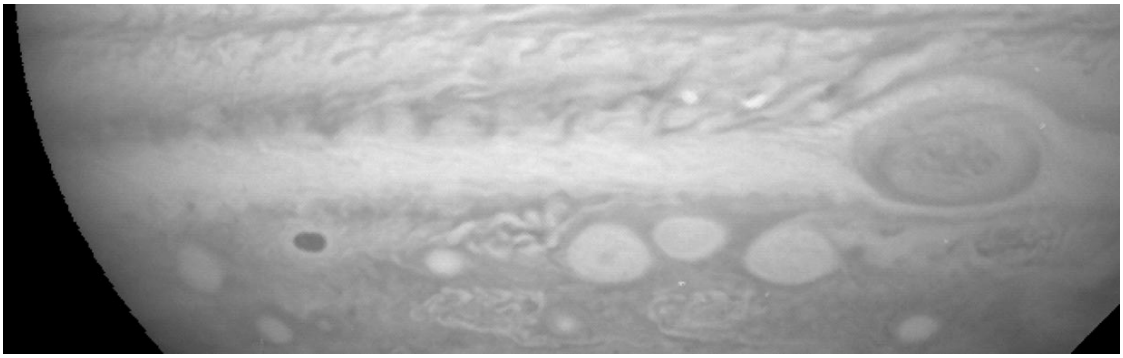

Fig. 3. The Great Red Spot in the Jovian atmosphere.

the coalescence of vortices driven by the Joukowski force [6]. Figure 5 illustrates the evolution of a Taylor–Couette flow from an initial state in which the "white" fluid (with zero "marker" concentration) in the annular half-gap adjoining the inner cylinder is at rest and the "black" fluid (with unit "marker" concentration) in the outer half-gap rotates as a solid body having the angular velocity of the outer cylinder. Here, the development of

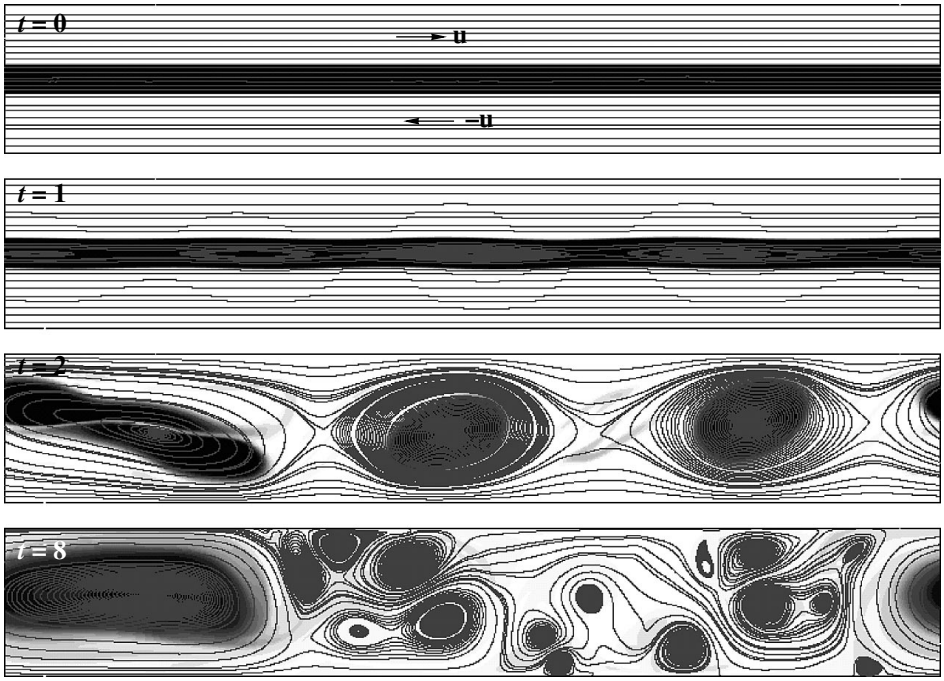

Fig. 4. Flow analogous to that shown in Fig. 1, but confined between walls.

the Kelvin–Helmholtz instability is followed by the survival of a few large vortices. Thus, the development of the intermediate spectral range can involve energy transfer from small scales as well.

Another important issue is local heat balance and the rate of energy redistribution in a local volume. The assumption of balanced energy transfer at every point of the flow is a very strong hypothesis that requires a special proof. When it is discarded, the turbulent energy spectrum must be computed by direct numerical simulation.

The results of time-dependent two-dimensional computations presented above provide evidence of both inverse energy transfer (as in two-dimensional turbulent flows) and energy transfer from large-scale vortices to small scales (as in Kolmogorov turbulence).

In [10], the linear stability problem was analyzed for the incompressible plane Couette flow with a free boundary. The free-boundary condition was set by prescribing a constant external pressure. It was found that the highest growth exponent is determined by the layer thickness, which agrees with the scenario of large-vortex formation simulated in our computations.

3. OUTLOOK

To summarize the foregoing discussion, we state two major problems that must be solved in order to develop a physical model of turbulence:

(1) generation of large-scale vortices by the pressure field in a shear flow at $\mathrm{Re} > \mathrm{Re}_{cr}$;

(2) development of a cascade of flow structures due to interaction between a shear flow and large-scale vortices.

Many open questions in turbulence theory will be answered after these problems are solved by direct numerical simulation.

One important issue is the scope of the Navier–Stokes equations. These equations, in particular, are known to describe the Couette flow of a viscous fluid. At present, it can be argued that the Navier–Stokes equations correctly describe viscous flows with Reynolds numbers below critical values. The reference length scale used in calculating the Reynolds number is determined by the characteristic length over which the Reynolds stress is comparable to the viscous stress in a shear flow. This length characterizes the large-scale structure of turbulent flow. Viscosity plays a key role in the conversion of kinetic energy into heat in small-scale turbulence and in the quasilaminar small-scale flow structure. (Note also that applied turbulence studies suggest that flows tend to quasilaminar regimes with increasing Reynolds number.) The minor role played by large-scale perturbation modes may be explained by strong interaction between vortices and the background flow. A more detailed explanation will possibly be inferred from direct numerical simulations to be performed in future. In recent studies of the behavior of viscosity coefficient in the Navier–Stokes equations, its complicated dependence on the velocity field was revealed (e.g., see [11]).

In [12], Troshkin introduced an infinite set of terms expressed in terms of higher order Lie derivatives of viscous terms into the Navier–Stokes equations, thus changing from linear to nonlinear viscosity. The new equations retain classical solutions (it was rigorously

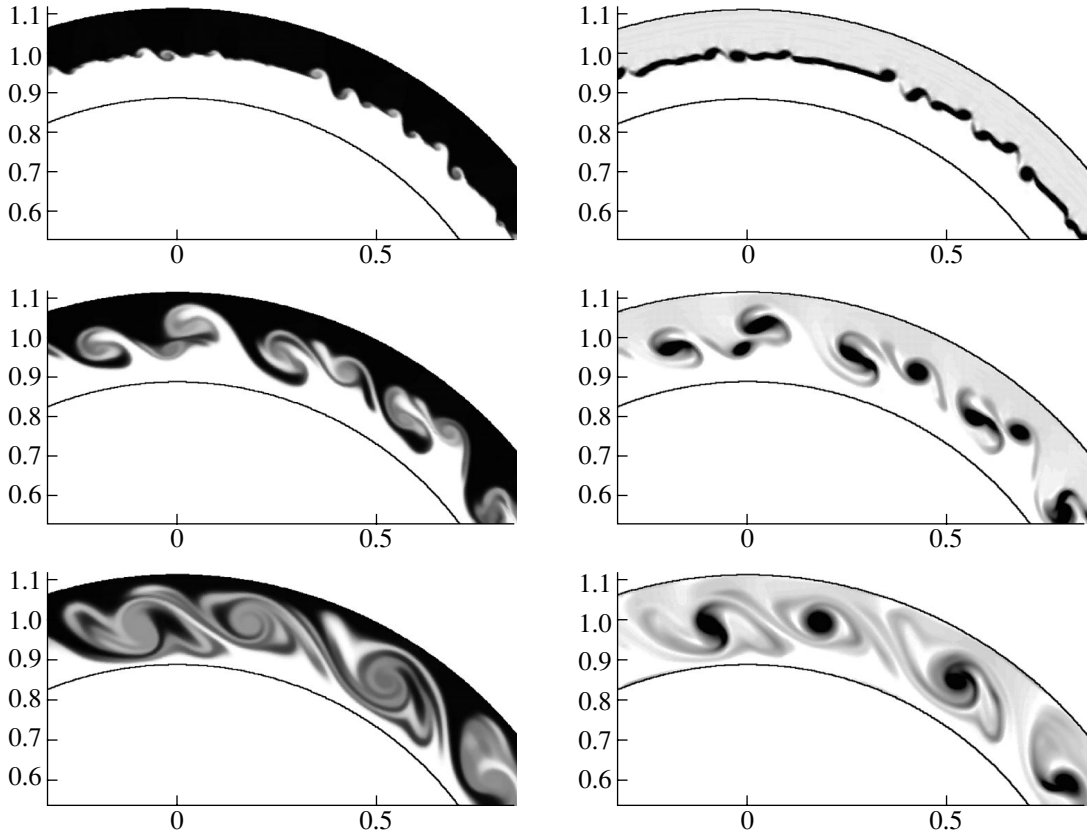

Fig. 5. "Marker" concentration (left) and vorticity (right) in the Taylor–Couette flow at $t = 0.05$, 0.1, and 0.2. The outer cylinder is rotating counterclockwise; the inner one is at rest.

proved that the Poiseuille flow structure remains invariant) and have new solutions, which may be of interest for turbulence theory. Physical reasons were put forward to explain why the analysis of turbulence based on linear viscosity (i.e., the classical Navier–Stokes equations) is inadequate. Thus, the analysis presented in [12] can be viewed as an attempt to expand the scope of conventional turbulence theories relying on numerical experiment, statistical methods, and averaged models by introducing generalized Navier–Stokes equations.

In [13], it was argued that energy input into a system at high Reynolds numbers [14] combined with the Kolmogorov law

$$\varepsilon \sim V^3 l^{-1} \text{ and } \nu_t \sim \varepsilon^{1/3} l^{4/3}$$

implies that

$$\mathrm{Re} = \frac{Vl}{\nu} \longrightarrow \frac{Vl}{\nu + \nu_t} \sim O(1).$$

In other words, turbulent viscosity is a varying parameter (and so is the Reynolds number). Here, V is interpreted as the velocity difference over a distance l. Therefore, turbulence will develop when the same value of V corresponds to a larger length scale (say, associated with pipe radius). In terms of energy, this means that the kinetic energy associated with a large length scale exceeds the work done by viscous forces on the corresponding volume. Similar reasoning can be used regarding the growth of velocity difference in a shear flow. These considerations suggest that the Navier–Stokes equations can be used to describe a viscous flow only when the Reynolds number is lower than a certain critical value. Analysis of compressible flows with higher velocities and greater length scales must rely on the Euler equations [3].

In view of the argumentation developed above, analysis of the onset of turbulence based on expansion in terms of small velocity perturbations does not seem to be well founded. It may be necessary to analyze developing turbulence by direct numerical simulation, since the problem is nonlinear and turbulence develops in a finite volume. Therefore, decomposition of flow velocity into an unperturbed mean component and a small disturbance is not valid since the velocity satisfies the equations of fluid dynamics at every point of the flow.

4. CONCLUSIONS

The numerical results presented in Figs. 1, 4, and 5 lead to the following conclusions about the processes underlying the onset of turbulence in a flow. The development of turbulence begins with formation of large vortices. The vortex size is determined by the length

scale on which the inertial term $\mathbf{v}\Delta\mathbf{v}$ is comparable to the viscous tensor. As a result, a vortex is created by a pressure gradient. We believe that representation of incipient turbulence in terms of small perturbations developing against the background velocity field is incorrect. The onset of turbulence is due to finite disturbances. Well-developed turbulence should be modeled in the framework of the Euler equations, which correctly describe the distribution of basic length scales [3]. This conclusion relies on the fact that the viscous term is small when the Reynolds number is high. When solving the Navier–Stokes equations, one must take into account the heating of the fluid due to kinetic-energy dissipation in small-scale turbulence.

The relationship between stochasticity and existence of large-scale structures in turbulent flows was discussed in [5]. In particular, it was emphasized that development of turbulence depends on imbalance between the inertial mechanism (responsible for growing flow instability) and viscous dissipation (which stabilizes the flow). We can introduce the concept of intrinsic inertial mechanism of a dynamical system, due to the nonlinear inertial terms in the governing equations and external perturbing forces (forces arising from the pressure field are also treated as external perturbations). Instability ("transition to chaos") can rapidly grow when the inertial mechanism dominates over dissipation. As noted in [5], the intrinsic inertial mechanism of a dynamical system may not be sufficient to initiate a stochastic process in a system.

Another important issue in turbulence theory is evolution of the energy spectrum. It is commonly believed that energy is transferred from large to small scales via eddy breakup (direct cascade). This assumption leads to the Kolmogorov–Obukhov spectrum of well-developed turbulence when the rate of spectral energy transfer is constant. According to our simulations of direct cascading, small- and intermediate-scale eddies appear as a result of interaction between the background flow and the very first large vortex generated in its wake.

Support for the proposed model can be found in [15], where the dominant role played by large-scale vortices was demonstrated by three-dimensional computations of shear flows.

Important results concerning the evolution of small-scale turbulence have been obtained. Entropy computations have shown that only a small fraction of the energy contained in small-scale eddies dissipates into heat [16]. Multiple small eddies merge under the action of the Joukowski force, and the resulting inverse cascade determines the inertial range spectrum.

Thus, the onset of turbulence includes the following stages. Large vortices develop at random locations in a free shear flow, with length scales as described above. Subsequently, interaction between the large vortices and the background flow gives rise to the generation of small-scale components of turbulence and determines its inertial-range spectrum. Thus, the onset of turbulence has a deterministic rather than stochastic nature. Only vortex locations are random, whereas the spectrum is controlled by interaction between the stream and large vortices.

Well-developed turbulence in a free shear flow involves both direct and inverse cascades, i.e., spectral energy transfer in both directions.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project nos. 03-01-00700 and 03-02-16548. The numerical simulations were performed under the program "Mathematical Modeling and Intellectual Systems."

REFERENCES

1. A. S. Monin and A. M. Yaglom, *Statistical Fluid Mechanics* (Nauka, Moscow, 1967; MIT Press, Cambridge, 1975), Part 2.
2. A. S. Monin, P. Ya. Polubarinova-Kochina, and V. I. Khlebnikov, *Cosmology, Hydrodynamics, and Turbulence* (Nauka, Moscow, 1989) [in Russian].
3. O. M. Belotserkovskiĭ, Zh. Vychisl. Mat. Mat. Fiz. **25**, 1856 (1985).
4. O. M. Belotserkovskii, in *Numerical Methods in Fluid Dynamics*, Ed. by H. J. Wirz and J. J. Smolderen (Hemisphere, Washington, 1978), p. 339.
5. O. M. Belotserkovskii, *Turbulence and Instabilities* (Lewinstoon, 2000).
6. O. M. Belotserkovskiĭ, A. M. Oparin, and V. M. Chechetkin, *Turbulence: New Approaches* (Nauka, Moscow, 2002) [in Russian].
7. A. S. Monin, in *Short Talks on Turbulence*, Ed. by O. M. Belotserkovskiĭ (Nauka, Moscow, 1994), p. 7 [in Russian].
8. J. P. Boris, F. F. Grinstein, E. S. Oran, and R. L. Kolbe, Fluid Dyn. Res. **10**, 192 (1992).
9. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics*, 3rd ed. (Nauka, Moscow, 1986; Pergamon, New York, 1987).
10. A. G. Doroshkevich, Astron. Zh. **57**, 259 (1980) [Sov. Astron. **24**, 152 (1980)].
11. V. V. Vedenyapin, *Bolzmann–Vlasov Kinetic Equations* (Fizmatlit, Moscow, 2001) [in Russian].
12. O. V. Troshkin, in *Mathematical Modeling: New Results*, Ed. by O. M. Belotserkovskiĭ and V. A. Gushchin (Nauka, Moscow, 2003) [in Russian].
13. V. M. Canuto, Mon. Not. R. Astron. Soc. **317**, 985 (2000).
14. H. Tennekes and J. L. Lumley, *A First Course in Turbulence* (MIT Press, Cambridge, 1972).
15. S. Yanase and Y. Kaga, in *Proceedings of Japan–Russia Seminar on Turbulence and Instabilities* (Tokyo, 2003).
16. O. M. Belotserkovskiĭ, A. M. Oparin, and V. M. Chechetkin, Zh. Vychisl. Mat. Mat. Fiz. **42**, 1727 (2002).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 3, 2004, pp. 510–521.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 3, 2004, pp. 585–599.

FLUIDS

Simulation of Mass Transfer in Systems with Isotropic Pair Correlation of Particles

O. S. Vaulina* and O. F. Petrov

Institute for High Energy Densities, Joint Institute for High Temperatures, Russian Academy of Sciences, Moscow, 127412 Russia

**e-mail: industpl@redline.ru*

Received February 4, 2004

Abstract—Results of a numerical analysis of mass transfer in systems of macroscopic particles with various isotropic interaction potentials are presented. Parameters that determine transport properties of nonideal dissipative systems are obtained for a broad class of model potentials. An approximate expression for the diffusivity of interacting particles is proposed. The relationship between diffusivity and viscosity is analyzed for strongly nonideal systems. *© 2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Problems concerning mass transfer in dissipative systems of interacting particles are of great interest for various branches of science and engineering (fluid dynamics, plasma physics, pharmaceutical industry, polymer physics and chemistry, etc.) [1–12]. Most difficulties in studies of physical properties of these systems arise from the fact that no analytical theory of liquids is available that can explain their thermodynamic properties, predict equations of state, describe heat and mass transfer, etc. [9–12]. Another problem is that real interparticle interaction potentials are not known in most cases, and the energy of interparticle interaction cannot be calculated exactly for real physical systems [11–17]. Currently, this problem is commonly considered to be unsolvable within the framework of molecular physics. For this reason, the standard approach used in modeling potential energy of particles relies on effective semiempirical potentials U_{eff}. These potentials are widely used in theoretical analyses of transport in real systems and studies of their physical properties by means of computer simulation. The most popular functions have the form

$$U_{\text{eff}} = -a/r^m + b/r^n$$

introduced by G. Mie about 100 years ago (a, b, n, and m are constants and r is the interparticle distance). The form of a model potential is almost always prescribed a priori, and the parameters of the chosen model function are adjusted by comparison with measured characteristics of the medium under analysis [15–17].

Modern theories of strongly nonideal systems and models describing the liquid state of matter are generally based on two approaches. In one approach, relations between different parameters of a liquid and their relations to the properties of corresponding crystals are determined by semiempirical methods based on analogies between the crystalline and liquid states of matter [9–12]. This approach was successfully applied in [18] to develop a hypothetical model of liquid structure at a temperature close to the melting point. Alternatively, the properties of a nonideal system are computed by the molecular dynamics method with the use of model data concerning the energy of interparticle interaction [16, 17].

Laboratory dusty plasmas provide good experimental models for studying properties of nonideal systems. A dusty plasma is a partially ionized gas containing micrometer-sized dust particles [19–23]. Analysis of the properties of dusty plasmas is also of interest because they are widespread in nature, on the one hand, and are frequently obtained in various technological processes, on the other. The micrometer-sized dust particles contained in these plasmas can acquire large electric charges and make up quasi-steady liquid- or solid-like dust structures. In contrast to real liquids, the macroscopic particles are sufficiently large to be imaged, which facilitates application of direct nonintrusive diagnostic methods and makes it possible to study kinetics of nonideal systems. The difference between dust-grain kinetics in a nonideal dusty plasma and kinetics of atoms (or molecules) in real molecular liquids is primarily due to difference in the interparticle interaction potential and strong influence of neutrals in the ambient gas on the motion of dust grains in a laboratory plasma.

However, the real potentials of interaction between dust particles in plasmas are not known either [24–27]. Therefore, determination of the parameters characterizing the state of a system of interacting particles is as important for the physics of nonideal dusty plasmas as it is for other natural sciences. In [28, 29], two dimensionless parameters were found to determine mass transfer and phase state for systems with Yukawa-type

screened Coulomb potentials (Yukawa systems with $\kappa = r_p/\lambda < 6$–7, where λ is the screening radius, $r_p = n_p^{-1/3}$ is the mean distance between particles, and n_p is the particle concentration): the effective nonideality parameter

$$\Gamma^* = \Gamma\{(1 + \kappa + \kappa^2/2)\exp(-\kappa)\}^{1/2}$$

and the scaling parameter

$$\xi = \nu_{fr}^{-1} e Z_p \{(1 + \kappa + \kappa^2/2)\exp(-\kappa) n_p/\pi m_p\}^{1/2}.$$

Here, Z_p is the charge of a dust particle, m_p is its mass,

$$\Gamma = \frac{(Z_p e)^2}{T r_p}$$

is the Coulomb coupling parameter, T is temperature, and ν_{fr} is the friction coefficient associated with collisions between dust particles and ambient-gas neutrals [30, 31]. Consistency of this numerical model (extended homogeneous system with a screened isotropic interaction between particles) with the conditions of laboratory experiments on gas-discharge dusty plasmas was verified in [32–34]. Experimental studies have shown that the dust-particle dynamics in these plasmas can be described in terms of the parameters Γ^* and ξ. However, the parameters of the interparticle interaction potential can be determined only if additional information about its form is available.

Thus, it was noted that mass transfer and spatial correlation of particles in Yukawa systems with $\kappa < 6$ depends only on the ratio of the second derivative U'' of the pair interaction potential $U(r)$ at r_p to the particle temperature T. Moreover, both melting/crystallization processes (at $\Gamma^* \sim 102$–106) and formation of well-ordered clusters of dust grains (at $\Gamma^* \sim 22$–25) occur at nearly constant values of Γ^* [28, 29]. One can reasonably assume that this property holds under certain conditions for potentials of more general form describing pair interactions. In the present study, we examine this hypothesis by computing transport coefficients and pair correlation functions for several models of radial repulsive potentials. The results obtained are compared with the diffusivities and pair correlation functions calculated for systems with attractive pair interaction potentials. We analyze available numerical data and measurement results concerning transport coefficients for elementary liquids (such as metals and inert gases).

2. PARAMETERS USED IN NUMERICAL ANALYSIS

Correct simulation of plasma-dust particle transport must rely on a molecular-dynamics method, in which a system of ordinary differential equations containing a Langevin force F_{br} is solved. This force represents random impacts by molecules of the ambient gas or other random processes that underlie the relaxation of the kinetic temperature T of dust grains to the equilibrium value characterizing the energy of their stochastic motion [7, 35, 36]. Microscopic processes in homogeneous extended clouds of interacting macroscopic particles are generally simulated by setting periodic boundary conditions and taking into account not only the random force F_{br}, but also the forces F_{int} acting between pairs of particles [28, 32–36]:

$$m_p \frac{d^2\mathbf{l}_k}{dt^2} = \sum_j F_{int}(l)\big|_{l=|\mathbf{l}_k-\mathbf{l}_j|} \frac{\mathbf{l}_k - \mathbf{l}_j}{|\mathbf{l}_k - \mathbf{l}_j|} - m_p \nu_{fr} \frac{d\mathbf{l}_k}{dt} + \mathbf{F}_{br}, \quad (1)$$

where $l = |\mathbf{l}_k - \mathbf{l}_j|$ is the separation between particles, and $F_{int} = -\partial U/\partial l$. Interaction between dust particles in plasmas is commonly described by the Yukawa-type screened Coulomb potential

$$U = a_0 \frac{\exp(-r/\lambda)}{r}, \quad (2)$$

where r is distance and a_0 is a parameter equal to $(eZ_p)^2$ for two identical particles with charge eZ_p (e is the elementary charge). This assumption is consistent both with measurements of forces acting between two dust particles [24] and with the computed structure of a screening cloud [25] only at relatively short distances from the particle ($r < 5\lambda_D$, where λ_D is the Debye radius). The screening weakens with increasing r, and the asymptotic behavior of U at $r \gg \lambda_D$ is governed by a power law: $U \propto r^{-2}$ [26] or $U \propto r^{-3}$ [27].

The results reported in [24–27] were obtained for isolated dust particles in plasmas. However, it remains unclear how the potential of interaction between two particles is modified by influence of other particles in a dust cloud, ionization of gas, collisions of electrons or ions with neutrals in the ambient gas, and other factors. Therefore, transport properties of nonideal systems were analyzed by using various combinations of power and exponential laws frequently employed in simulations of repulsion in kinetics of interacting particles [3–7]:

$$U = U_c a_1 (r_p/r)^n, \quad (3)$$

$$U = U_c[a_1 \exp(-\kappa_1 r/r_p) + a_2 \exp(-\kappa_2 r/r_p)], \quad (4)$$

$$U = U_c[a_1 \exp(-\kappa_1 r/r_p) + a_2 (r_p/r)^n], \quad (5)$$

where a_1, a_2, κ_1, κ_2, and n are parameters varied in our computations and $U_c = a_0/r$ is the Coulomb potential. Both potential (2) and models (4) and (5) (with $n = 1$ and 2 in the latter) are of special interest in the physics of dusty plasmas. They can be used to allow for weaker screening at relatively large distances between dust grains [24–27].

To analyze transport properties of systems of particles with various interaction potentials (2)–(5), three-

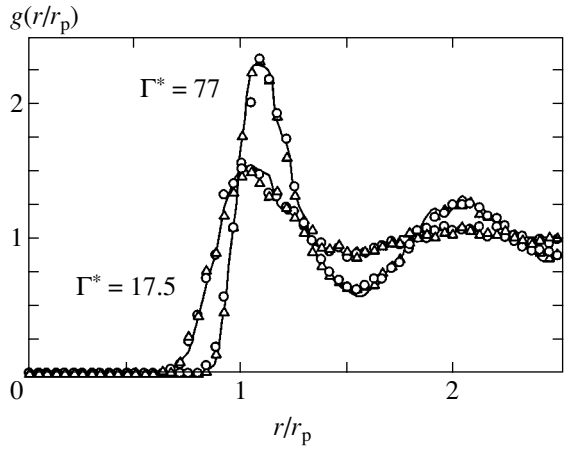

Fig. 1. Pair correlation functions $g(r/r_p)$ for several model potentials and several values of ξ and Γ^*. For $\Gamma^* = 77$, the solid curve, triangles, and circles correspond to $\xi = 0.14$ and $U/U_c = \exp(-4.8r/r_p)$, $\xi = 0.14$ and $U/U_c = 0.1\exp(-2.4r/r_p) + \exp(-4.8r/r_p)$, and $\xi = 1.22$ and $U/U_c = \exp(-4.8r/r_p) + 0.05r_p/r$, respectively. For $\Gamma^* = 17.5$, the solid curve, triangles, and circles correspond to $\xi = 1.22$ and $U/U_c = \exp(-2.4r/r_p)$, $\xi = 1.22$ and $U/U_c = 0.1\exp(-2.4r/r_p) + \exp(-4.8r/r_p)$, and $\xi = 0.14$ and $U/U_c = 0.05(r_p/r)^3$, respectively.

dimensional equations of motion (1) were solved for specific values of parameters defined by analogy with those characterizing Yukawa systems: the effective nonideality parameter

$$\Gamma^* = (Q_p^*)^2/Tr_p \tag{6}$$

and the scaling parameter

$$\xi = \omega^*/\nu_{fr}, \tag{7}$$

where the frequency of collisions between macroscopic particles is calculated as

$$\omega^* = Q_p^*(n_p/\pi m_p)^{1/2}, \tag{8}$$

and the effective particle charge is

$$Q_p^* = [U''/2n_p]^{1/2}. \tag{9}$$

Note that the effective particle charge does not have any particular physical meaning here. However, the use of (9) makes it possible to retain Γ^*, ξ, and ω^* as universally applicable parameters in models with interaction potentials of any type (2)–(5). In the case of Yukawa potential (2), the effective charge can be expressed as

$$Q_p^* = eZ_p(1 + \kappa + \kappa^2/2)^{1/2}\exp(-\kappa/2).$$

Most computations were performed for 125 independent particles in the central cell, while the number of particles taken into account in computing pair interactions reached approximately 3000. Yukawa potentials with $\kappa = 2.4$ and 4.8 were employed. Correct simulation of molecular dynamics was ensured by using discretization cells of size $R \gg \lambda$ [37]. The pair interaction potential was cut off at the distance $L_{cut} = 4r_p$. To ensure that numerical results are independent of number of particles and cut-off distance, we performed additional computations for 512 actual particles for $L_{cut} = 7r_p$ with $\Gamma^* = 1.5$, 17.5, 25, 49, and 92. The discrepancy between computed results did not exceed numerical error and remained within ±(1–3)%. A detailed description of the numerical procedure can be found in [28, 33, 34]. The value of ξ was varied between 0.04 and 3.6, i.e., within limits characteristic of experimental conditions in gas-discharge plasmas. The value of Γ^* was varied between 1 and 110.

Our computations showed that the effective parameter Γ^* completely characterizes mass transfer in the simulated systems, as well as their ordering and phase states, if the following empirical condition for long-range interaction is satisfied:

$$2\pi|U'(r_p)| > |U''(r_p)|r_p. \tag{10}$$

In the first (linear) approximation, this criterion specifies conditions under which the force acting between two particles separated by the mean interparticle distance is greater than the force typically arising in collisions of macroscopic particles (calculated by taking into account the probability of such collisions).

3. FORMATION OF ORDERED STRUCTURES

Ordering in the simulated systems was analyzed by using the pair correlation function $g(r)$ and the structure factor $S(q)$. Figure 1 compares these functions with the functions $g(r)$ computed by using various potentials subject to empirical condition (10) for two values of Γ^* and several values of ξ. Figure 2 illustrates the dependence of the first maxima g_1 and S_1 of the functions $g(r)$ and $S(q)$ on Γ^*. Here, vertical bars represent the absolute deviations of these quantities for $\xi \approx 0.04$–3.6 and potentials (2)–(5). To compare the pair correlation of particles computed for dissipative systems ($\nu_{fr} \neq 0$) with solutions to reversible equations of motion for nondissipative Yukawa systems ($\nu_{fr} = 0$, $\xi \longrightarrow \infty$) and with results obtained for a one-component plasma (OCP) model, Fig. 2 also shows the maximum values of $g(r)$ and $S(q)$ found in [38, 39].

Our numerical study shows that spatial correlation of dust grains in the simulated systems is independent of friction (ν_{fr}) and is completely determined by the value of Γ^* under conditions ranging from a gaseous state (at $\Gamma^* \approx 1$) to the point of crystallization into a BCC lattice (at $\Gamma^* \approx 102$–104). The first maximum points of the functions $g(r)$ and $S(q)$ corresponding to crystalline structure are characteristic of the BCC lattice:

$$d_{g1} \approx (3\sqrt{3}/4n_p)^{1/3}, \quad d_{S1} \approx 2\pi(\sqrt{2}n_p)^{1/3}.$$

Jumps in the values of the first maxima of $g(r)$ and $S(q)$ from 2.65 to 3.1 are observed as the normalized param-

eter Γ^* varies from the crystallization point $\Gamma_c^* \approx 102$–104 to the melting point $\Gamma_m^* \approx 106$–107 (see Fig. 2). Thus, $\Gamma_{cm}^* \approx 104.5(\pm 2)$ can be interpreted as the point of phase transition between a liquid-like state and a body-centered cubic lattice.

Since $\Gamma_{cm}^* \approx 104.5$ (±2%) is independent of the ambient viscosity, this result is consistent with molecular-dynamics simulations of crystallization in Yukawa systems with zero friction [38–41]. The deviations of their results from $\Gamma_{cm}^* \approx 104.5$ vary within ±5% and can be attributed to difference in numerical procedures (number of particles, integration step, etc.) and to choice of Γ^* associated with either melting or crystallization point of the system. It should be noted that $\Gamma_{cm}^* \approx 104.5$ (±2%) agrees with the theoretical results obtained in [42], where the value of the nonideality parameter on the phase-transition line in the OCP model was 105(±3%). (The latter value is consistent with numerical results based on various criteria for crystallization [43] and melting [44].)

Note also that the form of a correlation function $g(r)$ satisfying condition (10) is determined by the value of Γ^*. Therefore, the methods for determining the potential of interaction between two particles from measurements of the structure factor based on the hypernetted chain approximation [5, 6, 39] cannot be applied to the systems in question. Furthermore, the result obtained here can explain the widespread use of various phenomenological melting/crystallization criteria specifying the maximum values of correlation functions or the ratios of their maximum and minimum values on phase-transition lines (when $r \neq 0$) irrespective of the interparticle interaction potential.

4. TIME-DEPENDENT MASS TRANSFER

Unlike phase state, mass transfer in the simulated systems is determined by two parameters: Γ^* and ξ (see Figs. 3–5). To analyze the dust-grain diffusivity $D = \lim_{t\to\infty} D(t)$ and its time dependence $D(t)$, we used the expression

$$D(t) = \langle\langle \Delta l(t)^2 \rangle_N \rangle_t / 6t, \qquad (11)$$

where $\Delta l = \Delta l(t)$ is a particle displacement and angle brackets with subscripts N and t denote a quantity averaged over the ensemble of N particles and over all time intervals of duration t within the measurement time.

Figure 3 shows the ratio of diffusivity $D(t)$ of charged particles to the diffusivity $D_0 = T/\nu_{fr} m_p$ of noninteracting "Brownian" particles as a function of time (measured in units of ν_{fr}^{-1}) for several combinations of pair interaction potentials and values of ξ. Curve *6* is the exact solution to the Langevin equation in the absence of interaction between particles [7]:

$$D(t)/D_0 = 1 - \frac{1 - \exp(-\nu_{fr} t)}{\nu_{fr} t}. \qquad (12)$$

Thus, $D(t) = D_0$ for noninteracting particles at relatively

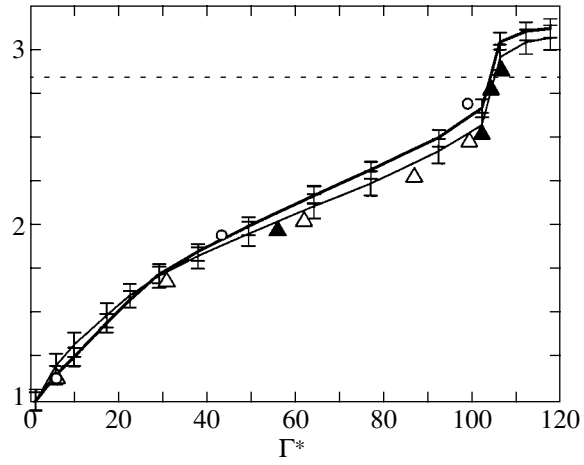

Fig. 2. First maxima S_1 of structure factor (thick curve) and g_1 of pair correlation function (thin curve) versus Γ^*: closed triangles represent g_1 in the nondissipative Yukawa model ($\nu_{fr} = 0$) [38]; open triangles, g_1 in the OCP model [39]; open circles, S_1 in the OCP model [39]. Vertical bars are the absolute deviations for $\xi = 0.04$–3.6 and for various potentials satisfying (10).

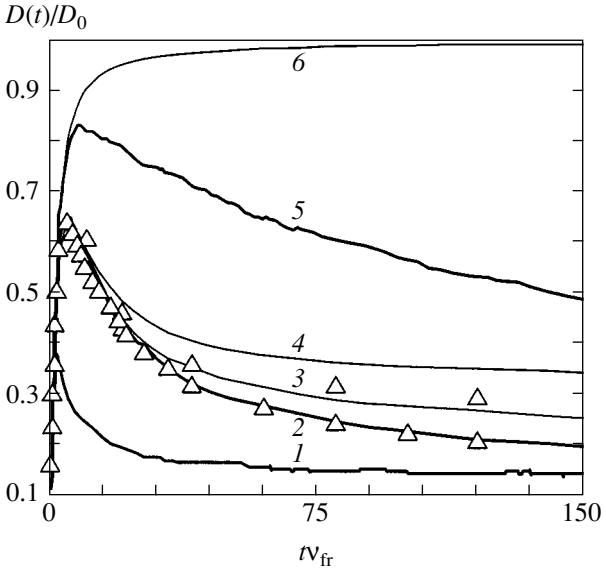

Fig. 3. Ratio $D(t)/D_0$ versus $t\nu_{fr}$ for the Yukawa potentials with $\kappa = 2.4$ for $\xi = 1.2$ and $\Gamma^* = 80$ (*1*), $\xi = 0.14$ and $\Gamma^* = 80$ (*2*), $\xi = 0.14$ and $\Gamma^* = 60$ (*3*), $\xi = 0.14$ and $\Gamma^* = 30$ (*4*), and $\xi = 0.04$ and $\Gamma^* = 80$ (*5*); curve *6* corresponds to Eq. (12). Triangles represent the data points obtained in computations with $\xi = 0.14$, $U/U_c = \exp(-4.8r/r_p) + 0.05 r_p/r$, and $\Gamma^* = 60$ and 80 corresponding to curves *3* and *2*, respectively.

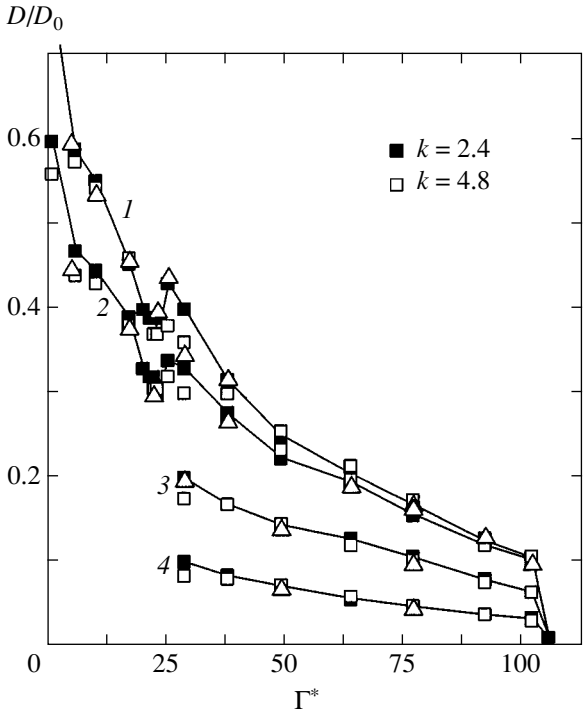

Fig. 4. Ratio D/D_0 versus Γ^* for dissipative Yukawa systems with different κ (■; □) [28] and for systems with potentials (2)–(4) specified in the caption to Fig. 1 (△) at ξ = 0.14 (*1*), 0.41 (*2*), 1.22 (*3*), and 3.65 (*4*).

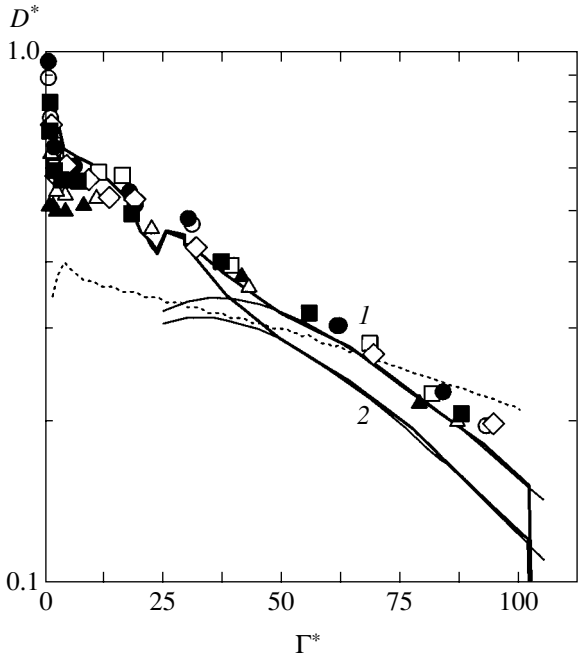

Fig. 5. Normalized function D^* averaged over numerical results versus Γ^* (thick curves); its approximation given by (16) for $\xi \geq 0.41$ (*1*) and $\xi \leq 0.14$ (*2*); and D^* for nondissipative Yukawa systems ($\xi \longrightarrow \infty$) [45] with κ = 0.16 (○), 0.48 (●), 0.97 (□), 1.61 (◇), 2.26 (■), 3.2 (△), 4.8 (▲), and 8 (dotted curve).

long times ($\nu_{fr}t \gg 1$), whereas ballistic motion of particles manifests itself at short times ($\nu_{fr}t \ll 1$): $\langle \Delta l(t)^2 \rangle \approx 3Tt^2/m_p$ and $D(t) \propto t$.

For interacting particles, similar short-time behavior of $D(t)$ is obtained and the ratio $D(t)/D_0$ is consistent with (12) at $t < t_{max}/2$ (see Fig. 3). The maximum value D_{max} reached by the function $D(t)$ can be used to analyze mass transfer at short times. Note that D_{max} approaches D_0 from below as the ambient viscosity is increased. Furthermore, both D_{max}/D_0 and $t_{max}\nu_{fr}$ are independent of Γ^*. An empirical approximation of the numerical results corresponding to the curves $D(t)$ that monotonically decrease to the left of their maxima gives the following expression for the function $D_{max}(\xi)$:

$$D_{max} \approx \frac{D_0}{1 + b\xi}. \tag{13}$$

For weakly dissipative systems with $\xi > 1$, the maximum point t_{max} of $D(t)$ is determined by the relation

$$t_{max}\nu_{fr} = \frac{\pi}{((b\xi)^2 - 1)^{1/2}}, \tag{14}$$

where $b = 4$. Formulas (13) and (14) agree with numerical results within 5%. They are obtained by solving the problem of one-dimensional oscillator with the frequency defined as

$$\omega_c^2 = \frac{2|U''(r_p)|}{\pi m_p} \equiv 4\omega^{*2}. \tag{15}$$

This frequency characterizes particle oscillation in a BCC lattice [28], and its value is close to the frequency of the average mode of particle oscillation for a liquid-like Yukawa system obtained by analyzing the corresponding particle-velocity autocorrelation functions [45]:

$$\omega_c^2 \approx \frac{2\omega_\perp^2 + \omega_\parallel^2}{4\pi},$$

where $\omega_\perp$ and $\omega_\parallel$ correspond to the transverse and longitudinal oscillation modes.

The variation of $D(t)$ simulated here illustrates relaxation of the simulated system for small deviations from an equilibrium state. Our computations suggest that a system of interacting particles can be characterized by equilibrium values of diffusivity only at $t \gg t_{max}$, whereas $D(t) = D_0$ for a system of "Brownian" particles at $t \gg \nu_{fr}^{-1}$. Analysis of the short-time behavior of $D(t)$ can be used both to estimate the parameter ξ as a source of information about the interparticle interaction potential and to determine the temperature T of dust grains if their velocity distribution cannot be measured with required accuracy.

5. TEMPERATURE-DEPENDENT DIFFUSIVITY

With time elapsed ($t \longrightarrow \infty$), the function $D(t)$ defined by (11) tends to its constant value $D = \lim_{t\to\infty} D(t)$ corresponding to the standard definition of diffusivity as a basic transport coefficient. Figure 4 shows the results of numerical simulations of dust-grain particles represented as the ratios of the diffusivity D of charged particles to the diffusivity D_0 of Brownian particles for different values of ξ and types of model potentials. It demonstrates that the diffusivities of particles with interaction potentials (2)–(5) are determined by two parameters: Γ^* and ξ. As the nonideality parameter approaches $\Gamma^* \approx 106$ and the system crystallizes into a BCC lattice (see above), the dust-grain diffusivity drops to zero (see Fig. 4).

The normalized function

$$D^* = \frac{D(1+\xi)}{D_0} \equiv \frac{D(\nu_{fr} + \omega^*)m_p}{T_p}$$

(averaged over $\xi \approx 0.045$–3.65 and potentials (2)–(5) used in the computations) is shown in Fig. 5. The deviations of this function calculated for different values of U and ξ do not exceed ±5% for all values of Γ^* except for the interval between 22 and 30, where the scatter of computed D^* is ±10% and D is a nonmonotonic function of Γ^* (see Figs. 4 and 5). The increase in scatter can be explained by formation of well-ordered clusters of dust grains at $\Gamma^* < 22$–25. This observation suggests that the type of interparticle interaction may affect the phase-transition dynamics. Formation of clusters of this type revealed in numerical studies of Yukawa systems [28, 46] supports the topological scenario of melting observed in some experiments [46–48].

Thus, we have found that the normalized function D^* is determined by Γ^* for both weakly nonideal and strongly correlated systems. In the latter case, the diffusivity for $\Gamma^* > 40$ can be represented by analogy with the diffusivity of macroscopic particles in Yukawa systems [28]:

$$D \approx \frac{T\Gamma^*}{12\pi(\omega^* + \nu_{fr})m_p}\exp\left(-c\frac{\Gamma^*}{\Gamma_c^*}\right), \qquad (16)$$

where $\Gamma_c^* = 102$ is the crystallization point, $c = 2.9$ for $\xi \geq 0.41$, and $c = 3.15$ for $\xi \leq 0.14$. The error of the approximation of computed D with the use of (16) does not exceed 3% for Γ^* between 50 and 102. The error increases to 7–13% as Γ^* decreases to 40 and reaches 25–30% for $\Gamma^* \approx 30$. Formula (16) agrees with the jump model developed for molecular liquids [1, 12] and can be used to calculate Γ^* by using the mean interparticle distances, temperatures, and particle diffusivities measured for liquid-like systems without introducing additional hypotheses about the pair interaction potential. Diagnostics of weakly correlated systems with $\Gamma^* < 50$ can rely on the numerically calculated D^* presented in Fig. 5. Previously, the influence of interparticle interaction forces on temperature-dependent diffusion was described by using either virial expansions or relations based on analogies with critical phenomena in liquids [2, 45, 49]. Both schemes required additional calculations of coefficients in these approximations when the system's parameters were varied.

To compare these results with diffusivities calculated for nondissipative Yukawa systems ($\nu_{fr} \equiv 0$, $\xi \longrightarrow \infty$), Fig. 5 shows the numerical results obtained in [45] for $\kappa \approx 0.15$–4.85 and $N_p = 600$–1000. It is clear that the normalized diffusivity D^* predicted for nondissipative systems agrees (within ±3%) with the results of simulations of weakly dissipative systems for $\xi \geq 0.41$ and Γ^* between 15 and 100. Unlike the present computational method (Brownian molecular dynamics under periodic boundary conditions), the approach used in [45] was based on solution of reversible equations of motion, which required renormalization of computed results to maintain thermodynamic equilibrium in the system, and an infinite system of particles was modeled by applying special tensor transformations. The latter approximation could lead to influence of different boundary conditions on the calculated diffusion coefficient and, therefore, certain disagreement (up to 10 to 15% for $\Gamma^* < 15$) between the numerical results shown in Fig. 5.

Finally, we note that increase in the screening parameter κ leads to failure of criterion (10) as applied to potential (2) at $\kappa > 5.1$. However, since this condition is empirical and sets a lower limit (π can be replaced with 4 in (10) for most potentials), the transport properties of Yukawa systems are determined by Γ^* and ξ for κ up to about 7. Figure 5 compares the value of D^* corresponding to $\kappa \approx 8$ [45] with the results obtained for $\kappa < 5$ and demonstrates an increasing difference in the behavior of diffusivity for Yukawa systems as the screening parameter is increased.

6. TRANSPORT PROPERTIES OF SYSTEMS WITH ATTRACTIVE PAIR INTERACTION POTENTIALS

Condition (10) for scaling of transport properties of nonideal systems with different interparticle interactions was obtained for isotropic repulsive potentials. Let us now analyze the transport properties of systems with effective potentials $U_{eff}(r) \equiv U(r)$ allowing for attraction between particles. As an example, we consider the pair correlation functions $g(r)$ and diffusivities D computed in [50] for the two potentials of this kind shown in Fig. 6. The potential represented by curve *1* is consistent with the results of measurements of $g(r)$ and diffusivity of sodium atoms at a temperature about 100°C. The potential represented by curve *2* agrees with the results computed for sodium by using the Born–Green theory [51]. Both potentials satisfy condition (10). In [50], the transport properties of sodium were computed by a molecular-dynamics method

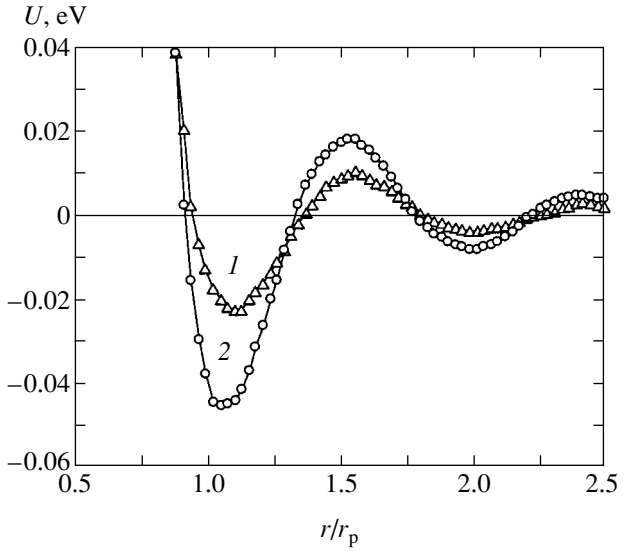

Fig. 6. Pair interaction potentials for sodium [50].

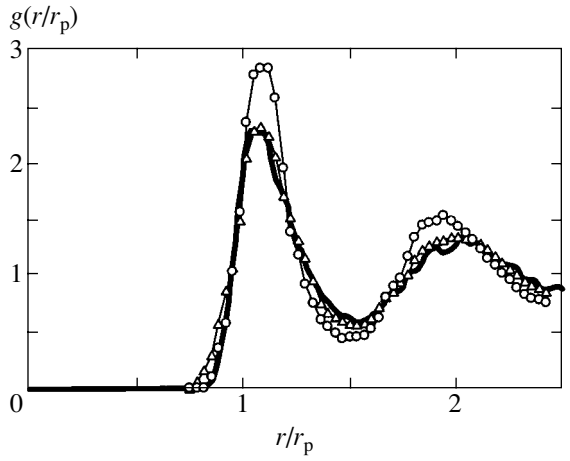

Fig. 7. Pair correlation functions $g(r/r_p)$ obtained for potentials *1* (△) and *2* (○) in Fig. 6 and pair correlation functions for a simulated system with Γ* = 77 (bold curve).

(based on reversible equations of motion) for about 600 particles. Both potentials were cut off at distances of about 8.2 Å and the mean interparticle distance was about 3.4 Å. Figures 7 and 8 show the results of computations of pair correlation functions and $D(t)$, respectively. For the potentials represented by curves *1* and *2* in Fig. 6, the values of $D = \lim_{t\to\infty} D(t)$ are approximately 5.8×10^{-5} and 1.9×10^{-5} cm^2/s.

For liquid sodium ($\nu_{fr} \equiv 0$), the diffusivity given by (16) can be expressed as

$$D \approx \frac{r_p}{12}\sqrt{\frac{T_c\Gamma_c^*}{\pi m_p}}\exp\left(-2.9\frac{\Gamma^*}{\Gamma_c^*}\right), \qquad (17)$$

where the crystallization point T_c almost coincides with the melting point T_m for the system. Since the pair correlation function computed for potential *1* in Fig. 6 is similar to our numerical results for $\Gamma^* \approx 77$, we used this value to determine the diffusivity D given by (17) for sodium atoms. The calculated result is $D \approx 5.77 \times 10^{-5}$ cm^2/s and is consistent with the numerical result reported in [50]. In view of the maximum $g_1 \approx 2.9$ of the correlation function obtained for potential *2* (see Fig. 7), it can be concluded that the state of the simulated system corresponds to the metastable interval between the melting and crystallization lines, $\Gamma^* \approx 104 \pm 2$ (see Fig. 2). Calculation of the diffusivity as given by (17) is not valid in this interval (see Fig. 4), and the predicted $D \approx 3 \times 10^{-5}$ cm^2/s is greater than $D \approx 1.9 \times 10^{-5}$ cm^2/s obtained in [50].

It should be noted that the maximum points t_{max} of the function $D(t)$ corresponding to the two potentials are consistent with the values of Γ* found by analyzing pair correlation functions. In particular, relation (14) predicts $\Gamma^* \approx 76.5$ ($t_{max} \approx 1.44 \times 10^{-13}$ s) and 106.5 ($t_{max} \approx 1.22 \times 10^{-13}$ s) for potentials *1* and *2*, respectively. However, lower values of the nonideality parameter were obtained by using expression (13) for D_{max} in both cases: $\Gamma^* \simeq 53.5$ ($D_{max} \simeq 7.4 \times 10^{-5}$ cm^2/s) and $\Gamma^* \simeq 84$ ($D_{max} \simeq 5.9 \times 10^{-5}$ cm^2/s), respectively. This may be explained by cutoff of pair interactions at short

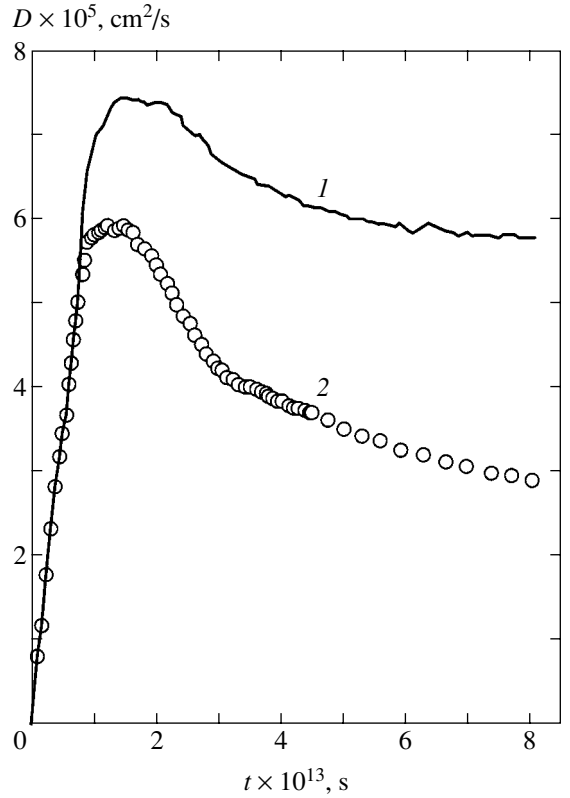

Fig. 8. Diffusivities $D(t)$ obtained for potentials *1* (△) and *2* (○) in Fig. 6.

distances (about $2.5r_p$) or an insufficiently large number of particles used in averaging the rms deviation Δl^2 [33].

It should also be noted that most available experimental results concerning the diffusivities for liquid metals near the corresponding melting lines agree with those given by (17) within measurement error. In particular, the diffusivity estimated for liquid sodium at $T \sim 100$°C by using available experimental data is $D \approx (3.8 \pm 0.6) \times 10^{-5}$ cm^2/s [18, 51–53], while expression (17) predicts $D(T = 100°\text{C}) \approx 3.3 \times 10^{-5}$ cm^2/s and $D(T = 125°\text{C}) \approx 4 \times 10^{-5}$ cm^2/s.

7. VISCOSITY OF WEAKLY DISSIPATIVE NONIDEAL SYSTEMS

An analysis of simulations performed for liquid sodium suggests that the numerical model considered here can be used to describe transport properties of real monatomic liquids that crystallize into BCC lattices (at least, liquid alkali metals). However, this hypothesis must be verified by invoking additional data on diffusivities (or correlation functions) in a sufficiently wide temperature range, which can hardly be found in reference books. Most available results of measurements of diffusivities for atoms or molecules in liquids are characterized by scatter comparable to the measured values, and the resulting correlation functions may strongly depend on the method employed [2, 51–54]. This is explained both by the systematic errors inherent in the techniques used to measure the transport properties for strongly nonideal systems (typically, x-ray or neutron scattering) and by the high sensitivity of the measured results to the purity of a sample, the method of its preparation, and temperature variations [2, 53]. A larger body of experimental data on viscosity of liquids is available, and the error of their measurements generally does not exceed 5% [53, 54]. However, viscosity is very difficult to evaluate theoretically, because of the lack of analytical models. Moreover, no analytical relation between the diffusivity and viscosity has been established even for a liquid consisting of nonrotating spherically symmetric molecules [1, 2, 7, 30]. Finding a relation between diffusivity and viscosity valid in a wide range of nonideality parameters is a challenging task in diagnostics of dusty plasmas as well. Data on the diffusivity of dust grains in such plasmas can be obtained by nonintrusive diagnostic methods, such as processing of experimental video images and analysis of rms displacements of dust grains as functions of time [32–34]. Available methods for measuring viscosity involve dynamic perturbation of the system under study, which may affect the properties of a self-consistent plasma-dust system.

In the present study, we sought an empirical approximation of the shear viscosity η by using the numerical values of η calculated on the basis of the Yukawa model (with κ varying between 0.16 and 4.8) [55] and the OCP-lattice model [56, 57] by various authors in the limit of vanishing viscosity ($\xi \longrightarrow \infty$, $\nu_{fr} \longrightarrow 0$). A relation between the dust-grain diffusivity D and η was determined by invoking the well-known Stokes–Einstein formula (valid at low Reynolds numbers, $\mathrm{Re} \ll 1$), which is usually applied to describe diffusion of large molecules in a solvent with low molecular mass when a relation between transport coefficients is sought for a metal in the vicinity of the melting line [10, 18]:

$$\eta \approx \frac{T}{6\pi D a_{\mathrm{eff}}}. \tag{18}$$

Here, the effective radius a_{eff} of a spherical molecule is known up to a factor of 2 even in the case of large molecules [53].

The numerical data were approximated by using two procedures. One of them is based on the assumption that

$$a_{\mathrm{eff}}(T) \approx r_p - a_{\mathrm{WS}} \equiv \mathrm{const},$$

where a_{WS} is the Wigner–Seitz radius. Combined with (18), it yields the following relation between transport coefficients:

$$\eta \approx \frac{T}{7.5 D r_p}. \tag{19}$$

The other approximation relies on the following temperature dependence of the effective particle radius:

$$a_{\mathrm{eff}}(T) = (T/T_c)^{1/2} r_p/\pi,$$

which reflects the increase in the effective radius with T and leads to

$$\eta \approx \frac{\sqrt{T T_c}}{6 D r_p}. \tag{20}$$

Taking into account the numerical results concerning D and proposed temperature dependence (17), we find that both (19) and (20) evaluated for $\Gamma^* > 40$ agree with Andrade's semiempirical formula

$$\eta \propto f(T) \exp\left(\frac{W}{T}\right),$$

where W is the activation energy for self-diffusion and $f(T)$ is a weaker function of temperature as compared to exponential. This formula, based on the aforementioned jump model, provides a key analytical relation used to approximate the temperature dependence of viscosity for a liquid only within relatively narrow intervals of temperature variation.

To illustrate the discrepancy between proposed approximations (19) and (20) and the simulation results of [55–57], we normalized the viscosity η to

$$\eta_0 = \frac{\Gamma^*}{r_p^2}\sqrt{\frac{\pi T m_p}{\Gamma_c^*}}.$$

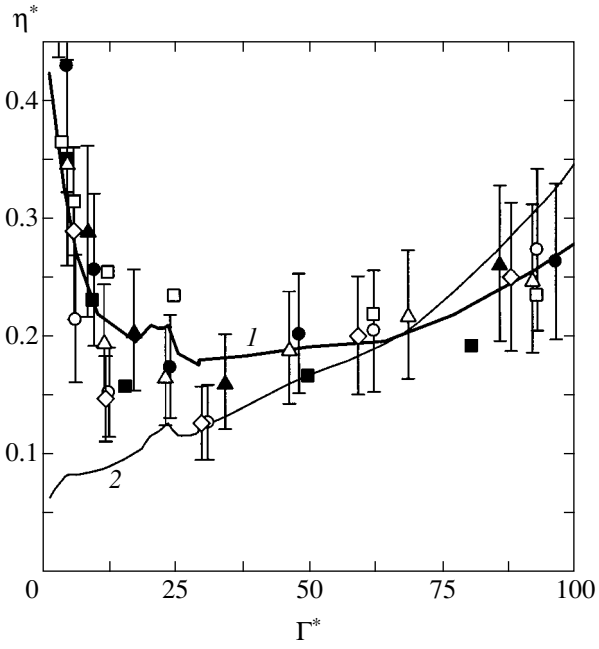

Fig. 9. Normalized viscosity η* versus Γ* predicted by (19) (*1*); (20) (*2*); BCC model for nondissipative systems in [56] (■) and [57] (□); and Yukawa model ($\xi \longrightarrow \infty$) [55] with $\kappa = 0.16$, (○), 0.81 (◇), 1.61 (●), 3.2 (△), and 4.8 (▲). Vertical bars represent confidence intervals corresponding to relative numerical errors of 20%.

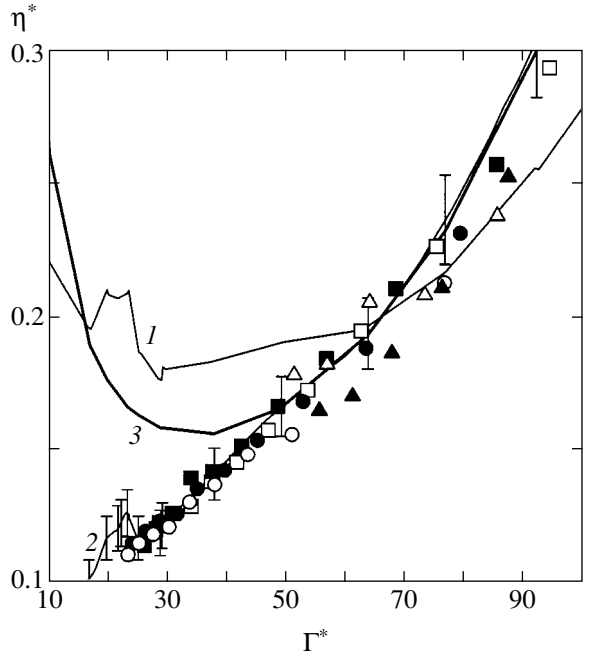

Fig. 10. Normalized viscosity η* versus Γ* predicted by (19) (*1*), (20) (*2*), and (21) (*3*). The function $\eta^*(\Gamma^* = \Gamma_c^* T_c/T)$ is shown for several liquid metals [53]: potassium, BCC (■); sodium, BCC (□); ruthenium, BCC (●); cesium, BCC (○); lead, FCC (▲); tin, TET (△). Vertical bars represent confidence intervals corresponding to relative numerical errors of 5% in (20).

This was done to improve graphic presentation of results and numerical errors, because the values of viscosity obtained for $\Gamma^* \approx 0.5$–100 varied by about an order of magnitude. Figure 9 shows the normalized viscosities $\eta^* = \eta/\eta_0$ obtained in all cases under analysis. It is clear that (19) (based on the assumption of constant effective particle radius) provides the best approximation of all available numerical data in the entire interval $\Gamma^* \approx 1$ and 100. However, since the results under analysis are characterized by large errors, both approximations lead to rms deviations within "numerical noise" for strongly correlated systems with Γ* between 20 and 100 (the rms errors of viscosity computations mentioned by the authors are at least 20%).

Figure 10 shows the experimental values of viscosities for various liquid metals borrowed from [53] and normalized to η_0. We used the mean interparticle distances r_p corresponding to the solid-phase densities,

$$r_p = r_{pc}, \quad \eta_0 = \frac{\Gamma^*}{r_{pc}^2}\sqrt{\frac{\pi T m_p}{\Gamma_c^*}},$$

and defined the nonideality parameter as

$$\Gamma^* = \Gamma_c^* T_c/T.$$

It is clear that approximation (20) allowing for temperature dependence of the effective particle size leads to very good agreement between measured viscosities and calculated particle diffusivities. The discrepancy between the experimental results and the predictions based on (20) does not exceed 3% for alkali metals (which crystallize into BCC lattices) and 10% for tin and lead, which have tetragonal (TET) and face-centered (FCC) lattices, respectively. Since the diffusivities for strongly correlated liquids with $\Gamma^* > 40$ can be represented by (17), the shear viscosity of the liquid metals considered here can be expressed analytically as follows in the temperature interval between T_c and $2T_c$:

$$\eta \approx \frac{2}{r_{pc}^2}\sqrt{\frac{\pi T m_p}{102}}\exp\left(2.9\frac{T_c}{T}\right). \qquad (21)$$

This approximation agrees with experimental data within the error of viscosity measurements. The results obtained demonstrate the high accuracy of approximations (20) and (21) of viscosity for the liquid metals considered here. Therefore, the corresponding diffusivities can be determined much more accurately than before both from available experimental data on η (by using (20)) and directly from numerical results (see Fig. 5 and formula (17)). Note also that (20), unlike the approximation based on $a_{eff}(T)$ = const proposed in [18], is valid not only on the melting line.

For comparison, let us consider the viscosities of liquid dielectrics crystallizing into FCC lattices (such as liquefied inert gases). An analysis of the viscosities of neon and argon measured on the corresponding saturation lines [53, 54] taking into account the density

variation with temperature and pressure, $\rho = m_p n_p$, shows that the relations

$$r_p = r_{pc}, \quad \frac{\Gamma^*}{\Gamma_c^*} = \frac{\rho T_c}{\rho_c T},$$

where ρ_c and ρ denote the solid- and liquid-phase densities, can be used in (20) to calculate the viscosities of these liquids up to an error of 5 to 7% in the entire temperature interval between the melting point and the critical point shown by the dashed line in Fig. 11. The normalized viscosities η^* of liquid neon and argon are also plotted as functions of

$$\Gamma^* = \Gamma_c^* \frac{\rho T_c}{\rho_c T}$$

in Fig. 11. Using expression (17) for diffusivity with $r_p = r_{pc}$, the viscosity of a liquefied inert gas in the temperature interval between T_c and $2\rho T_c/\rho_c$ can be represented as

$$\eta \approx \frac{2}{r_{pc}^2}\sqrt{\frac{\pi T m_p}{102}}\exp\left(2.9\frac{T_c\rho}{T\rho_c}\right). \qquad (22)$$

However, the good approximation of viscosity obtained here is not sufficient to conclude that model (20) is correct and the numerical results can be used to analyze the diffusivity of atoms in liquefied inert gases. Such a conclusion can rely only on experimental data concerning the value of the diffusivity or the qualitative behavior of $D(T)$.

Two remarkable facts should also be mentioned here. First, the behavior of the viscosities of liquefied inert gases and metals characterized by different lattice types in their solid states is in good agreement with the behavior of the transport coefficients obtained in simulations of nonideal systems crystallizing into BCC lattices, which are characterized by lower density. In the case of an inert gas, this can be explained by the fact that the coordination number Z changes in the transition from crystalline to liquid state from $Z = 12$ to $Z = 8.5$, which is close to the values of Z for metals with BCC lattices. However, the reduction of coordination number on the melting line cannot explain analogous behavior of transport coefficients for tin and lead, because it is believed that both packing of atoms and atomic bonds are preserved in solid–liquid phase transitions. Second, both the critical nonideality parameter

$$\Gamma_{cr}^* = 102 T_c/T_{cr} \approx 15$$

for alkali metals and the critical parameter

$$\Gamma_{cr}^* = 102(\rho_{cr}T_c)/(\rho_c T_{cr}) \approx 18$$

are close to $\Gamma^* \approx 22.5$ (see Figs. 4 and 5), which corresponds to formation of well-ordered clusters of dust grains [27, 46].

Correlation of the critical parameters and solid-phase parameters for materials having similar chemical properties is one of the methods used to determine their characteristics or physical properties in the vicinity of phase-transition lines when experimental data are not available [15, 17, 58]. Such materials are frequently considered as thermodynamically similar, and the similarity is attributed to similarity of effective potentials. For monatomic substances, the correlation of critical and solid-phase parameters manifests itself in similar ratios of T_{cr} and T_m, as well as those of the critical-state density ρ_{cr} and the solid-phase density ρ_c, depending on the position of the element in the periodic table. In particular, $T_{cr}/T_m = 6.755 \pm 0.015$ and $\rho_c/\rho_{cr} = 4.44 \pm 0.03$ for alkali metals [53, 54]. For inert gases, $T_{cr}/T_m = 1.81 \pm 0.01$ and $\rho_c/\rho_{cr} = 3.115 \pm 0.005$; for elements in the sixth group (e.g., Se or S), $T_{cr}/T_m = 3.30 \pm 0.05$ and $\rho_c/\rho_{cr} = 3.904 \pm 0.001$ [53, 54]. A correlation of the melting point, valence, and Fermi energy for simple densely packed metals was noted in [51], where it was shown that there is no simple relation between the sublimation (vaporization) temperature of the substance and the valence of the element. However, the critical point correlates with valence, because phase transition at this point is global, and the buildup of a new phase is not related to surface energy, being completely determined by interparticle interaction characteristic of the entire system. This suggests that the mean interatomic forces F in liquid metals are virtually independent of temperature. In view of the behavior of viscosity, it follows that $F(T, \rho) \approx$ const for alkali metals and $F(T, \rho) \propto \rho$ for liquefied inert gases. This observation can be used to construct model equations of state for such liquids.

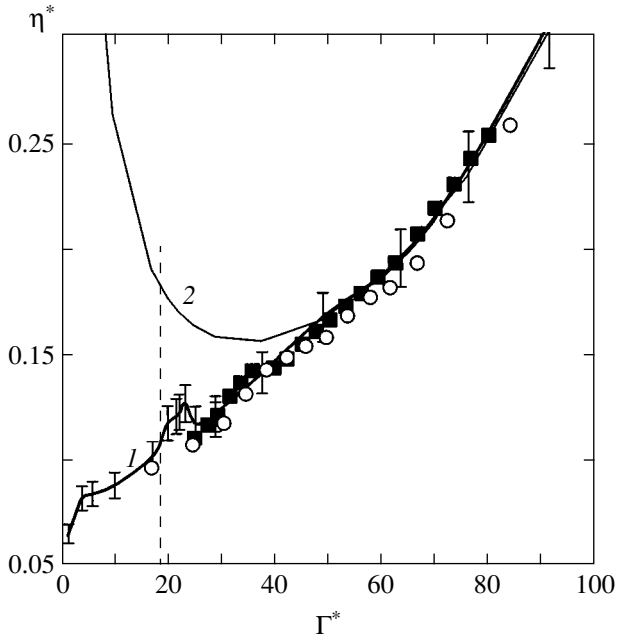

Fig. 11. Normalized viscosity η^* versus Γ^* predicted by (20) (*1*) and (22) (*2*). The function $\eta^*(\Gamma^* = \Gamma_c^* \rho T_c/(\rho T))$ is shown for liquid dielectrics [53, 54]: neon (■) and argon (○).

8. CONCLUSIONS

We have performed a numerical analysis of spatial correlation, mass transfer, and diffusion of dust grains with interparticle interaction potentials of different types for a wide range of phase states of nonideal systems ($\Gamma^* \approx 1$–110). We have found dimensionless parameters that determine the phase state and transport properties of dissipative systems with long-range potentials satisfying empirical condition (10): the effective nonideality parameter Γ^* and the scaling parameter ξ. In combination with temperature, these parameters determine the transport properties of the systems in question. It has been shown that the pair correlation function $g(r)$ depends on Γ^*, being virtually independent of friction forces (represented by ξ) for a broad class of isotropic potentials. This finding rules out any possibility of determination of these potentials by inverting $g(r)$.

Furthermore, we have found an analytical approximation of the diffusivity D of particles in strongly nonideal systems as a function of particle temperature T and the dimensionless parameters Γ^* and ξ. We propose a formula relating diffusivity to viscosity for elementary liquids (such as liquid metals and monatomic gases), which makes it possible to determine the corresponding diffusivities much more accurately than before. For dusty plasmas, the diffusivity of dust grains should be related to the dust-subsystem viscosity by conducting appropriate experiments. Moreover, these experiments could provide an answer to a fundamental question: are dusty plasmas analogs of real liquids (such as liquid metals) or systems having unique transport properties? Currently available experimental data on rotation velocities of dust grains in vortices suggest that the approximate relation between transport coefficients based on the assumption of independence of the effective particle radius of particle temperature should be preferred [59].

Data on diffusivity and viscosity are required to analyze the onset of various instabilities and dynamics of large-scale transport processes, such as formation of vortices or shock waves. The results presented here can be used to develop new nonintrusive diagnostic methods and examine the influence of interparticle interaction on transport properties under actual experimental conditions.

ACKNOWLEDGMENTS

This work was supported, in part, by the Russian Foundation for Basic Research, project no. 04-02-16362, and INTAS, grant no. 01-0391.

REFERENCES

1. Ya. I. Frenkel', *Kinetic Theory of Liquid* (Nauka, Leningrad, 1975; Oxford Univ. Press, Oxford, 1976).
2. *Photon Correlation and Light Beating Spectroscopy*, Ed. by H. Z. Cummins and E. R. Pike (Plenum, New York, 1974; Mir, Moscow, 1978).
3. B. Pullman, *Intermolecular Interactions: From Diatomics to Biopolymers* (Wiley, Chichester, 1978).
4. I. M. Torrens, *Interatomic Potentials* (Academic, New York, 1972).
5. R. Balescu, *Equilibrium and Nonequilibrium Statistical Mechanics* (Wiley, Chichester, 1975; Mir, Moscow, 1978).
6. N. K. Ailawadi, Phys. Rep. **57**, 241 (1980).
7. A. A. Ovchinnikov, S. F. Timashev, and A. A. Belyĭ, *Kinetics of Diffusely-Controlled Chemical Processes* (Khimiya, Moscow, 1986) [in Russian].
8. H. M. Thomas and G. E. Morfill, Nature **379**, 806 (1996).
9. A. A. Zhukhovitskiĭ and L. A. Shvatsman, *Physical Chemistry* (Metallurgiya, Moscow, 1987) [in Russian].
10. A. P. Regel' and V. M. Glazov, *Regularities in Formation of the Structure of Electronic Melts* (Nauka, Moscow, 1982) [in Russian].
11. D. K. Belashchenko, *Transport Phenomena in Liquid Metals and Semiconductors* (Atomizdat, Moscow, 1970) [in Russian].
12. Ya. I. Frenkel', *Introduction to the Theory of Metals* (Nauka, Leningrad, 1972) [in Russian].
13. I. G. Kaplan, *Theory of Intermolecular Interactions* (Nauka, Moscow, 1982; Elsevier, New York, 1986).
14. E. Mason and T. Sperling, *The Virial Equation of State* (Pergamon, New York, 1969; Mir, Moscow, 1972).
15. V. L. Patrushev, Zh. Éksp. Teor. Fiz. **8** (10–11), 1221 (1938).
16. G. G. Petrik, Z. R. Gadzhieva, and B. E. Todorovskiĭ, *Chemistry and Computer Simulation. Butlerov Lectures, Application to Special Issue*, No. 10, 301 (2002).
17. D. A. Young and B. J. Alder, Phys. Rev. A **3**, 364 (1971).
18. A. V. Gorshkov, Prikl. Fiz., No. 6, 65 (1999).
19. J. Chu and L. I, Phys. Rev. Lett. **72**, 4009 (1994).
20. H. Thomas, G. Morfill, V. Demmel, *et al.*, Phys. Rev. Lett. **73**, 652 (1994).
21. A. Melzer, T. Trottenberg, and A. Piel, Phys. Lett. A **191**, 301 (1994).
22. V. E. Fortov, A. P. Nefedov, V. M. Torchinskiĭ, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **64**, 86 (1996) [JETP Lett. **64**, 92 (1996)].
23. A. M. Lipaev, V. I. Molotkov, A. P. Nefedov, *et al.*, Zh. Éksp. Teor. Fiz. **112**, 2030 (1997) [JETP **85**, 1110 (1997)].
24. U. Konopka, G. Morfill, and R. Ratke, Phys. Rev. Lett. **84**, 891 (2000).
25. J. E. Daugherty, R. K. Porteous, M. D. Kilgore, *et al.*, J. Appl. Phys. **72**, 3934 (1992).
26. J. E. Allen, Phys. Scr. **45**, 497 (1992).
27. D. Montgomery, G. Joyce, and R. Sugihara, Plasma Phys. **10**, 681 (1968).
28. O. S. Vaulina and S. V. Vladimirov, Plasma Phys. **9**, 835 (2002).
29. O. S. Vaulina, S. V. Vladimirov, O. F. Petrov, *et al.*, Phys. Rev. Lett. **88**, 245002 (2002).

30. E. M. Lifshitz and L. P. Pitaevskiĭ, *Physical Kinetics* (Nauka, Moscow, 1979; Pergamon Press, Oxford, 1981).
31. N. A. Fuchs, *The Mechanics of Aerosols* (Dover, New York, 1964).
32. V. E. Fortov, O. S. Vaulina, O. F. Petrov, *et al.*, Phys. Rev. Lett. **90**, 245005 (2003).
33. O. S. Vaulina, O. F. Petrov, V. E. Fortov, *et al.*, Fiz. Plazmy **29**, 698 (2003) [Plasma Phys. Rep. **29**, 642 (2003)].
34. V. E. Fortov, O. S. Vaulina, O. F. Petrov, *et al.*, Zh. Éksp. Teor. Fiz. **123**, 798 (2003) [JETP **96**, 704 (2003)].
35. V. V. Zhakhovskiĭ, V. I. Molotkov, A. P. Nefedov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **66**, 392 (1997) [JETP Lett. **66**, 419 (1997)].
36. O. S. Vaulina, A. P. Nefedov, O. F. Petrov, *et al.*, Zh. Éksp. Teor. Fiz. **118**, 1319 (2000) [JETP **91**, 1147 (2000)].
37. R. T. Farouki and S. Hamaguchi, Appl. Phys. Lett. **61**, 2973 (1992).
38. M. O. Robbins, K. Kremer, and G. S. Grest, J. Chem. Phys. **88**, 3286 (1988).
39. S. Ichimaru, Rev. Mod. Phys. **54**, 1017 (1982).
40. E. J. Meijer and D. Frenkel, J. Chem. Phys. **94**, 2269 (1991).
41. S. Hamaguchi, R. T. Farouki, and D. H. E. Dubin, Phys. Rev. E **56**, 4671 (1997).
42. W. L. Slattery, G. D. Doollen, and H. E. DeWitt, Phys. Rev. A **21**, 2087 (1980).
43. E. L. Plolloc and J. P. Hansen, Phys. Rev. A **8**, 3110 (1973).
44. H. M. Van Horn, Phys. Lett. A **28A**, 707 (1969).
45. H. Ohta and S. Hamaguchi, Phys. Plasmas **7**, 4506 (2000).
46. O. S. Vaulina, O. F. Petrov, V. E. Fortov, *et al.*, Zh. Éksp. Teor. Fiz. **125**, 584 (2004) [JETP **98**, 515 (2004)].
47. G. Morfill, H. Thomas, and M. Zuzic, in *Advances in Dusty Plasma*, Ed. by P. K. Shukla, D. A. Mendis, and T. Desai (World Sci., Singapore, 1997), p. 99.
48. W. T. Juan, M. H. Chen, and L. I, Phys. Rev. E **64**, 016402 (2001).
49. J. M. A. Hofman, Y. J. H. Clercx, and P. J. M. Schram, Phys. Rev. E **62**, 8212 (2000).
50. A. Parskin and A. Rahman, Phys. Rev. Lett. **16**, 300 (1966).
51. *Physics of Simple Liquids*, Ed. by H. N. V. Temperley, J. S. Rowlinson, and G. S. Rushbrooke (North-Holland, Amsterdam, 1968).
52. P. D. Randolph, Phys. Rev. **134**, A1238 (1964).
53. *Physical Quantities. Handbook*, Ed. by I. S. Grigor'ev and E. Z. Meĭlikhov (Énergoatomizdat, Moscow, 1991) [in Russian].
54. N. B. Vargaftik, *Handbook on the Thermophysical Properties of Gases and Liquids* (Fizmatgiz, Moscow, 1963) [in Russian].
55. T. Saigo and S. Hamaguchi, Phys. Plasmas **9**, 1210 (2002).
56. J. Wallenborn and M. Baus, Phys. Rev. A **18**, 1737 (1978).
57. Z. Donko and B. Nyiri, Phys. Plasmas **7**, 45 (2000).
58. A. S. Basin, Izv. Sib. Otd. Akad. Nauk SSSR, Ser. Tekh. Nauk, No. 8 (2), 122 (1972).
59. O. S. Vaulina, O. F. Petrov, V. E. Fortov, *et al.*, Phys. Scr. (2004) (in press).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 3, 2004, pp. 522–529.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 3, 2004, pp. 600–608.

SOLIDS
Structure

Atomic Mechanisms of Pure Iron Vitrification

A. V. Evteev*, A. T. Kosilov, and E. V. Levchenko
Voronezh State Technical University, Voronezh, 394026 Russia
**e-mail: evteev@vmail.ru*
Received February 13, 2004

Abstract—It is shown on the basis of the model of iron with the Pak–Doyama paired potential of interatomic interaction in the framework of the molecular dynamics method that structural stabilization of the amorphous phase of pure iron during hardening from melt is ensured by the formation of a percolation cluster from mutually penetrating and contacting icosahedrons with atoms at vertices and centers. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The establishment of basic differences between the atomic structures of melt and metallic glass is one of the most urgent and unsolved problems in the physics of the amorphous state. This is due to the fact that available diffraction methods for studying the structure of disordered materials provide only a one-dimensional averaged pattern of atomic distribution in the form of structural functions [1], while the proposed models of short-range order on the basis of a random packing of atoms (e.g., the Bernal model) or coordination polyhedrons (e.g., the model based on Frank–Casper polyhedrons) can be applied equally successfully for describing the mutual arrangement of atoms in liquid as well as amorphous metals [2–5], but individual features of their structural organization remain unclear.

Considerable progress in the formation of model concepts concerning the structure of liquid and amorphous states of metallic systems, as well as atomic mechanisms of their mutual transformations, can be achieved by carrying out computer experiments in the framework of the molecular dynamics method using highly effective well-optimized algorithms and appropriate atomic interaction potentials.

A detailed study of the structure of condensed media requires analysis of coordination polyhedrons formed by the atoms of the system and formulation of geometrical laws underlying their mutual spatial arrangement. In the case of crystals, this problem is substantially simplified since translation invariance reduces the volume under investigation to a unit cell in which atoms occupy fixed positions, being located at certain distances from one another. For structurally disordered condensed media, such as melts and amorphous solids, the absence of translation invariance creates a complex mosaic of local atomic configurations, which can be studied correctly and systematically only by constructing a computer model of the structure of the object under investigation and analyzing coordination polyhedrons formed by all atoms of the structure.

The results of statistical-geometrical and cluster analysis of the structure of the molecular dynamic model of iron with the Pak–Doyama paired interatomic potential [6] in the course of hardening from melt, which will be considered below, are a continuation of the model concepts formulated in [7] and concerning the structure of pure amorphous metals and the atomic mechanisms of its formation.

2. DESCRIPTION OF THE MODEL

First, a molecular-dynamic model of liquid iron at $T = 2300$ K with a density of 7800 kg/m^3 was constructed (the density was specified in accordance with the data obtained for α-Fe [8] taking into account a correction of ~1% for amorphization). For the initial structure, we used a random dense packing of atoms. The interatomic interaction was described with the help of the empirical Pak–Doyama paired potential, which has the form [6]

$$\begin{aligned}\phi(r) &= -0.188917(r - 1.82709)^4 \\ &+ 1.70192(r - 2.50849)^2 - 0.198294 \text{ eV},\end{aligned} \quad (1)$$

where r is expressed in angstroms. The potential cutoff radius (the distance at which the potential and its first derivative smoothly vanish) is $r_c = 3.44$ Å. The parameters of this potential are determined from the data on elastic properties of α-Fe. The application of this potential in simulating liquid and amorphous iron [9–11] and its alloys with metalloids [12, 13] ensures good agreement between the calculated and experimental structural characteristics. The model contained 100000 atoms in the main cube with periodic boundary conditions. At the initial instant, the velocities corresponding to the Maxwell distribution were imparted to atoms. The method of calculations based on the molecular dynamics method involved numerical integration of the equation of motion with a time step of $\Delta t = 1.523 \times 10^{-15}$ s in

accordance with the Verlet algorithm [14]. The system was held at the above temperature for 3000 time steps (isothermal conditions). The temperature constraint was suspended and thermal equilibrium stabilized in the system over 3000 time steps at a constant internal energy (adiabatic conditions).

Then the system was isochorically cooled at a rate of 4.4×10^{12} K/s. The cyclic procedure of hardening was reduced to a stepwise decrease in temperature by ΔT = 20 K ($T_i = (115 - i)\Delta T$, where T_i is the temperature of the "surrounding medium" in the ith cycle), maintaining this temperature in the system over $1000\Delta t$, followed by annealing under adiabatic conditions for $2000\Delta t$. Thus, the duration of a cycle was $3000\Delta t$ or 4.569×10^{-12} s. Averaging of the thermodynamic parameters of the system (temperature T, potential energy U, and the product of pressure and volume PV) was carried out during the last $1000\Delta t$ of each cycle. It should be noted that the temperature T of the system under adiabatic conditions and the temperature T_i of the surrounding medium are not strictly identical. After each cycle, the system was transformed to a state with $T = 0$ by the static relaxation method, allowing the atoms to occupy the equilibrium positions in local potential wells. For statically relaxed models, the potential energy U_0 and the product of pressure by volume P_0V were calculated and the structural analysis was carried out. Such a procedure made it possible to determine the extent of structural relaxation of the model upon cooling.

3. RESULTS AND DISCUSSION

Figure 1 shows that the potential energy and the product of pressure by volume change continuously upon cooling, but the derivatives of these quantities at temperature $T_i \sim 1180$ K of the surrounding medium decrease sharply. The temperature dependences of the potential energy and the product of pressure and volume exhibit after static relaxation a point of inflection at the same temperature, which also follows from the peak on the derivatives of these dependences (Fig. 1). It was also found that the mean total displacement of atoms decreases sharply at $T_i < 1180$ K, which is characteristic of the change in the mechanism of atomic displacement from the activationless (for diffusion of atoms in a liquid) to the activation mechanism (for diffusion of atoms in a solid) [15].

The statistical-geometrical analysis of the structure of molecular-dynamic models of liquid and amorphous iron on the basis of Voronoi polyhedrons [2–5] proved that 70 and 80% of all atomic configurations, respectively, can be described by Voronoi polyhedrons with different combinations of tetra-, penta-, and hexagonal faces only. More than 90% of all atoms in the models of both liquid and amorphous iron have a geometrical coordination number from 12 to 15. It should be recalled that an individual Voronoi polyhedron can be described by a set of n_q numbers equal to the number of faces with q sides (n_3–n_4–n_5–…) [2–5]. This set of numbers can also be used for describing the coordination polyhedrons. In this case, n_q indicates the number of vertices at which q edges converge. The results led to 28 topologically allowed coordination polyhedrons of the type (0–n_4–n_5–n_6) with coordination numbers from 12 to 15.

If we denote the numbers of vertices, edges, and faces in a Voronoi polyhedron by V_V, E_V, and F_V, respectively, these quantities satisfy, in accordance with the Euler theorem, the following relation:

$$V_V - E_V + F_V = 2. \tag{2}$$

Since the intersection of three planes is sufficient for the formation of a vertex of a Voronoi polyhedron, and the probability of the formation of a vertex at the intersection of four and more planes is too low in view of geometrical distortions in the amorphous structure, the following equality holds for a Voronoi polyhedron of atoms in the amorphous structure:

$$2E_V = 3V_V. \tag{3}$$

This equality indicates that each edge connects two vertices and three edges meet at each vertex. For a coordination polyhedron, the numbers of vertices, edges, and faces are $V_C = F_V$, $E_C = E_V$, and $F_C = V_V$, respectively; hence, it follows from relation (3) that coordination polyhedrons constructed for atoms of an amorphous structure can have only triangular faces.

Taking into account relation (3), we can write the Euler equation (2) in the form

$$3F_V - E_V = 6. \tag{4}$$

For a Voronoi polyhedron of the type (0–n_4–n_5–n_6), using Eq. (4) with allowance for

$$F_V = n_4 + n_5 + n_6, \tag{5}$$

$$E_V = \frac{1}{2}(4n_4 + 5n_5 + 6N_6), \tag{6}$$

we obtain the relation

$$n_5 = 2(6 - n_4). \tag{7}$$

Since n_4 and n_5 must be positive integers, expression (7) indicates that their values are bounded and connected pairwise. The numerical value of n_6 for each such pair and for a preset coordination number from the interval $12 \le F_V \le 15$ can be determined using relation (5).

The table contains 28 topologically allowed coordination polyhedrons of the type (0–n_4–n_5–n_6) satisfying the condition $12 \le n_4 + n_5 + n_6 \le 15$ as well as their

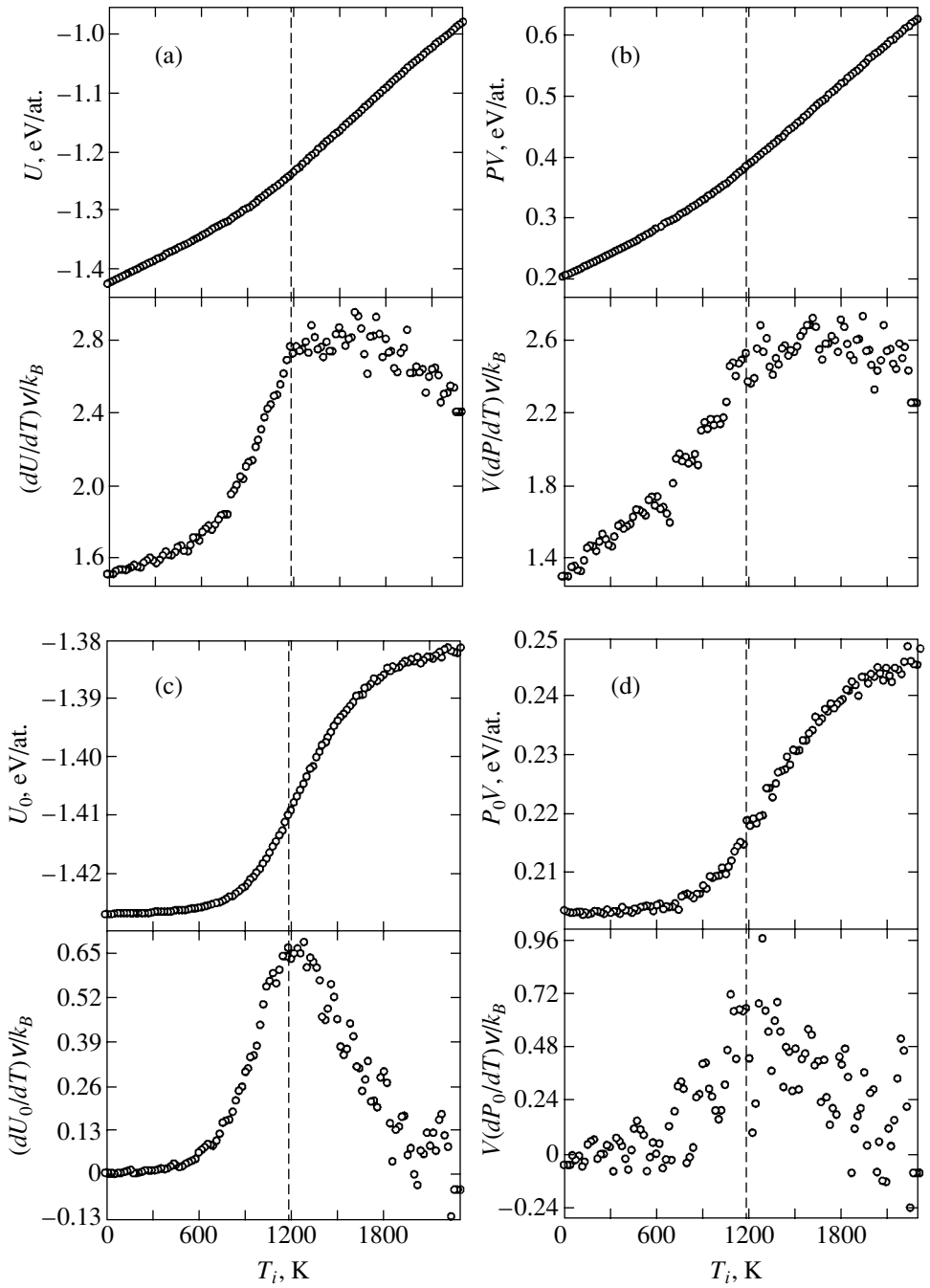

Fig. 1. Dependence of the thermodynamic function in the model of iron and their derivatives on the temperature T_i of the surrounding medium under cooling from the melt at a rate of 4.4×10^{12} K/s: (a) potential energy U; (b) product of pressure by volume, PV; (c) potential energy after static relaxation, U_0; (d) product of pressure by volume after static relaxation, P_0V.

molar fraction in the models of liquid and amorphous iron at the "ambient" temperatures of 2300 and 0 K. Fourteen main types of coordination polyhedrons (their fraction is greater than 1%) for the liquid and amorphous states are shown by bold face. The remaining coordination polyhedrons are thermodynamically unstable.

It can be seen from the table that, in the process of hardening from melt, the most intense increase (~4%) is observed in the fraction of atoms for which the coordination polyhedron (0–0–12–0) is an icosahedron and the Voronoi polyhedron is a dodecahedron (Fig. 2). It should be noted that icosahedron is the most compact and stable from the energy point of view among all coordination polyhedrons encountered in densely packed structures (both ordered and disordered). For example, the energy of an isolated icosahedron (coordi-

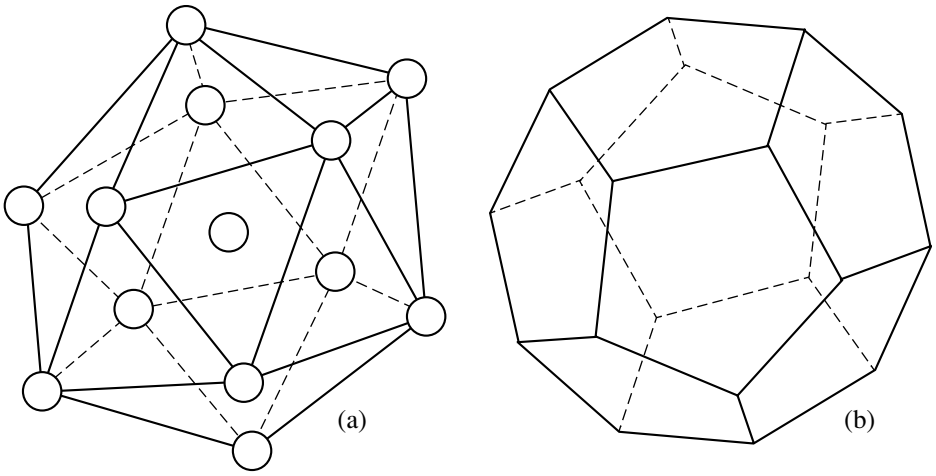

Fig. 2. (a) Coordination polyhedron (icosahedron) and (b) its Voronoi polyhedron (dodecahedron).

nation number 12) calculated using the Pak–Doyama paired potential amounts to –0.79 eV/at. For comparison, the energies of isolated coordination polyhedrons in hcp and bcc lattices (cubooctahedron with a coordination number of 12 and a rhombododecahedron with a coordination number of 14) amount respectively to –0.69 and –0.73 eV/at. However, the existence of pentad axes in an icosahedron makes it incompatible with translational symmetry; consequently, the formation of an amorphous structure of pure metals on the basis of such coordination polyhedrons hampers the process of atomic rearrangement during crystallization, creating prerequisites for stabilization of the amorphous state. For this reason, we mainly studied regularities in the evolution of structural elements with icosahedral symmetry during hardening.

Distribution of topologically allowed coordination polyhedrons of the type (0–n_4–n_5–n_6) with coordination numbers satisfying the condition $12 \le V_C \le 15$ in the model of iron in the liquid and amorphous states. The first number indicates the fraction of coordination polyhedrons in the melt at T_i = 2300 K, the second, in the amorphous state after hardening at T_i = 0. Bold face indicates coordination polyhedrons whose fraction is greater than 1% both in the melt and in the amorphous state

V_C	12	13	14	15
(0–0–12–n_6)	**(0–0–12–0)**	(0–0–12–1)	**(0–0–12–2)**	**(0–0–12–3)**
Fraction, %	4.315 ⟶ 8.033	0	2.180 ⟶ 4.933	1.218 ⟶ 2.589
(0–1–10–n_6)	(0–1–10–1)	**(0–1–10–2)**	**(0–1–10–3)**	**(0–1–10–4)**
Fraction, %	0	10.144 ⟶ 13.183	4.882 ⟶ 7.105	3.569 ⟶ 4.820
(0–2–8–n_6)	**(0–2–8–2)**	**(0–2–8–3)**	**(0–2–8–4)**	**(0–2–8–5)**
Fraction, %	3.001 ⟶ 1.961	2.521 ⟶ 1.620	9.783 ⟶ 10.981	3.962 ⟶ 4.432
(0–3–6–n_6)	(0–3–6–3)	**(0–3–6–4)**	**(0–3–6–5)**	**(0–3–6–6)**
Fraction, %	1.137 ⟶ 0.391	9.867 ⟶ 9.215	5.337 ⟶ 4.284	2.617 ⟶ 2.076
(0–4–4–n_6)	(0–4–4–4)	(0–4–4–5)	**(0–4–4–6)**	(0–4–4–7)
Fraction, %	0.433 ⟶ 0.131	1.108 ⟶ 0.548	3.940 ⟶ 3.068	1.149 ⟶ 0.796
(0–5–2–n_6)	(0–5–2–5)	(0–5–2–6)	(0–5–2–7)	(0–5–2–8)
Fraction, %	0.044 ⟶ 0.003	0.469 ⟶ 0.259	0.006 ⟶ 0	0.159 ⟶ 0.088
(0–6–0–n_6)	(0–6–0–6)	(0–6–0–7)	(0–6–0–8)	(0–6–0–9)
Fraction, %	0.001 ⟶ 0	0.002 ⟶ 0	0.158 ⟶ 0.252	0.001 ⟶ 0
Total, %	8.754 ⟶ 10.519	24.111 ⟶ 24.825	26.286 ⟶ 30.623	12.675 ⟶ 14.801

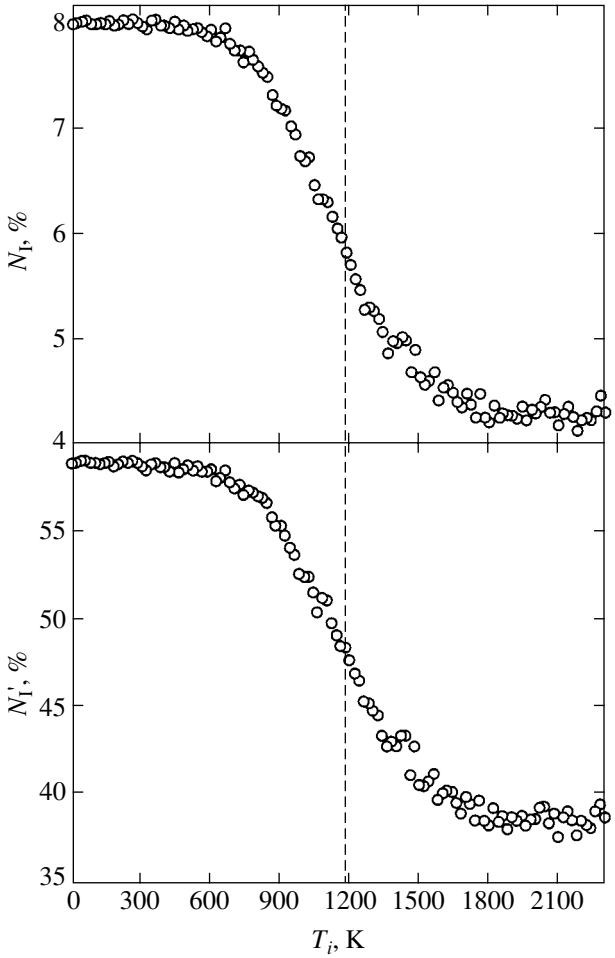

Fig. 3. Dependence of the number of atoms N_I located at the centers of icosahedrons and the total number of atoms N_I' involved in the construction of icosahedrons on the temperature T_i of the surrounding medium. The cooling rate is 4.4×10^{12} K/s.

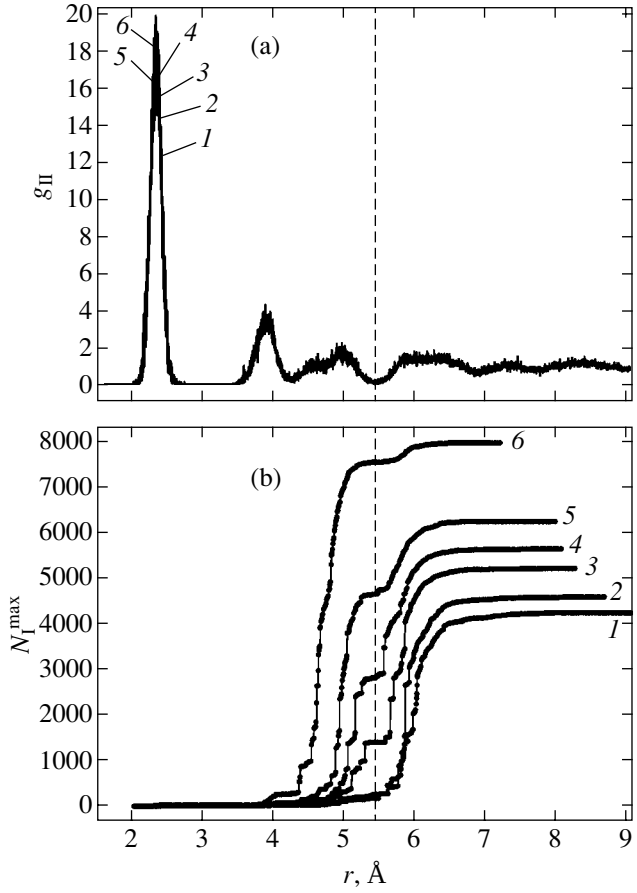

Fig. 4. (a) Paired function $g_{II}(r)$ of the radial distribution of atoms located at the centers of icosahedrons and (b) the number N_I^{max} of such atoms in the largest cluster with distances between neighbors smaller than or equal to r. The dashed vertical line separates the regions of mutually penetrating and contacting (left) and isolated (right) icosahedrons. Figures *1*, *2*, *3*, *4*, *5*, and *6* correspond to temperatures of 2300, 1500, 1300, 1200, 1100, and 0 K of the surrounding medium. The cooling rate is 4.4×10^{12} K/s.

As follows from Fig. 3, the fraction of atoms located at the centers of icosahedrons in the initial model of the liquid is 4.315%, while the total fraction of atoms participating in the construction of icosahedrons is ~38.813%. Each icosahedron contains on the average ~9 atoms. As the temperature decreases during hardening, the number of icosahedrons increases, and the above fractions in the most structurally relaxed model at 0 K attain values of 8.033% and 59.024%, respectively. Each icosahedron contains on the average ~7 atoms belonging simultaneously to several icosahedrons; in other words, the fraction of mutually penetrating icosahedrons contacting with one another increases. The points of inflection on these curves, which correspond to the maximal rate of increase in the number of icosahedrons, correspond to $T_i \sim 1180$ K, as on the curves describing the temperature dependences of the potential energy and the product of pressure by volume after static relaxation.

It is interesting to note that the number of atoms located at the centers of icosahedrons free of mutually penetrating contacts remains virtually unchanged during hardening, while the number of pairs of mutually penetrating icosahedrons substantially increases.

Figure 4a shows paired functions $g_{II}(r)$ of the radial distribution of atoms located at the centers of icosahedrons, which were calculated at temperatures of 2300, 1500, 1300, 1200, 1100, and 0 K of the surrounding medium. It should be noted that the form of these functions is virtually independent of temperature. The first peak on the $g_{II}(r)$ curve corresponds to the contact of mutually penetrating icosahedrons, the second peak, to the contact of icosahedrons at faces and edges, and the

third (split) peak, to the contact of icosahedrons at vertices.

To study the regularities in the structural organization of icosahedral coordination polyhedrons upon cooling of the model of iron in the framework of the percolation theory, we carried out cluster analysis. Figure 4b shows the number $N_{\mathrm{I}}^{\max}$ of icosahedrons in the largest cluster with distances between neighboring atoms smaller than or equal to r as a function of r for temperatures of 2300, 1500, 1300, 1200, 1100, and 0 K of the surrounding medium. It can be seen from the figure that, at $T_i > 1200$ K, the percolation threshold is observed to the right of the dashed vertical line separating the region of mutually penetrating icosahedrons contacting one another from the region of isolated icosahedrons. Consequently, at $T_i > 1200$ K, a percolation cluster of mutually penetrating and contacting clusters is not formed. Direct observations of the behavior of atoms located at the centers of icosahedrons revealed that isolated finite clusters of mutually penetrating icosahedrons contacting with one another is formed in this case; these icosahedrons continually change their shape and size with decreasing temperature. As the temperature decreases in the course of hardening below 1200 K, the percolation threshold is shifted to the left of the dashed vertical line (see Fig. 4b) to the region of mutually penetrating and contacting icosahedrons. In this case, a percolation cluster consisting only of mutually penetrating and contacting icosahedrons is formed; the cluster subsequently grows and pierces the entire structure.

Figure 5 shows the change in the size of the largest cluster of mutually penetrating icosahedrons contacting with one another upon cooling during hardening. The distances between neighboring atoms in the icosahedrons are smaller than or equal to $r = 5.46$ Å (position of the vertical dashed line in Fig. 4b). Figure 6 shows the projections of such a cluster onto a face of the computational cell at temperatures of 1460, 1300, 1260, 1200, 1180, and 0 K of the surrounding medium. It can be seen from Figs. 5 and 6 that, at $T_i < 1460$ K, fluctuations of the size of the largest cluster of mutually penetrating and contacting icosahedrons sharply increase; at $T_i < 1180$ K, the maximal growth of the cluster is observed (due to merging of coarse clusters), while at $T_i > 1180$ K, a monotonic increase takes place due to joining of small-size clusters and single icosahedrons.

These regularities in the self-organization of the icosahedral substructure during vitrification correlate well with the temperature dependences of the basic thermodynamic parameters (see Fig. 1), which exhibit some features of a second-order phase transition [16] and explain their behavior on microscopic level. Periodic boundary conditions do not allow us to exactly determine the glass formation temperature T_{g} (the temperature at which a percolation cluster is formed from mutually penetrating and contacting icosahedrons) of

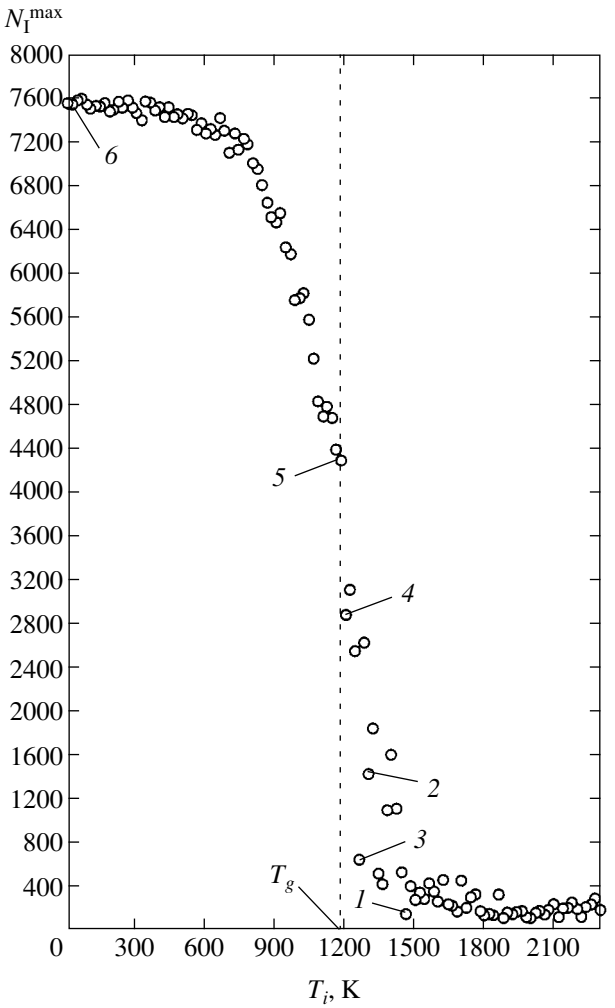

Fig. 5. Dependence of the size $N_{\mathrm{I}}^{\max}$ of the largest cluster formed by mutually penetrating and contacting icosahedrons on the temperature T_i of the surrounding medium. The vertical dashed line corresponds to the glass formation temperature T_{g}. Figures *1*, *2*, *3*, *4*, *5*, and *6* correspond to temperatures of 1460, 1300, 1260, 1200, 1180, and 0 K of the surrounding medium and to the numbers of cluster projections in Fig. 6. The cooling rate is 4.4×10^{12} K/s.

the model since the percolation cluster is essentially an infinitely large structure incommensurate with the computational cell. We can only speak of a certain interval in which the glass formation temperature lies. For example, for the model of iron under consideration, the glass formation temperature lies in the interval $T_{\mathrm{g}} = 1180 \pm 80$ K (see Fig. 5). However, we can intuitively expect that an increase in the size of the computational cell will hamper the formation of a percolation cluster and, hence, the glass formation point will be shifter towards lower temperatures. In the model of iron at 0 K, the largest cluster has a fractal dimension of $D = 2.99$ and consists of 7559 mutually penetrating and contacting icosahedrons, whose construction involves

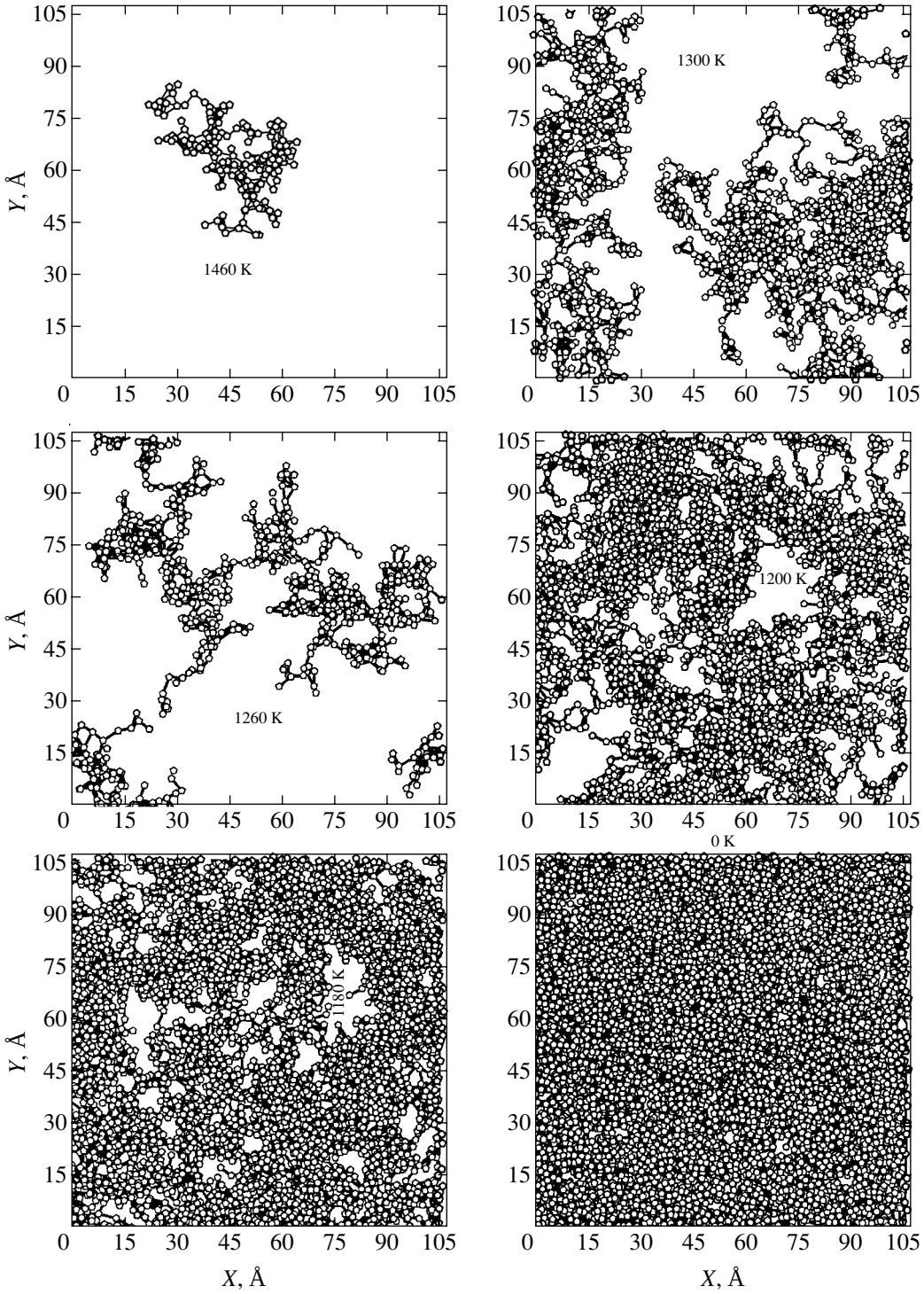

Fig. 6. Projections of the largest cluster formed by mutually penetrating and contacting icosahedrons onto one of the faces of the computational cell at various temperatures of the surrounding medium: 1460 (*1*), 1300 (*2*), 1260 (*3*), 1200 (*4*), 1180 (*5*), and 0 K (*6*).

54.485% of all atoms in the model. When recalculated to 1 mm^3, this value decreases to approximately 52.5%.

A decrease in the cooling rate in the model of iron melt to 6.6×10^{11} K/s leads to crystallization [17, 18].

4. CONCLUSIONS

The results considered above can be formulated in the following basic concept clarifying the physical nature of vitrification of pure iron.

Structural organization of the amorphous phase of pure iron is based on the formation (at the glass formation point) and subsequent growth of a percolation cluster of mutually penetrating icosahedrons coming in contact with one another, which contain atoms at the vertices and at the centers. A fractal cluster of icosahedrons incompatible with translational symmetry (the formation of these icosahedrons involves more than half of all the atoms in the system) plays the role of binding carcass hampering crystallization and serves as the fundamental basis of structural organization of the solid amorphous state of iron, which basically distinguishes it from the melt.

This statement introduces for the first time the concept of the substructure of amorphous metals.

REFERENCES

1. A. F. Skryshevskiĭ, *Structural Analysis of Liquids and Amorphous Solids* (Vysshaya Shkola, Moscow, 1980) [in Russian].
2. J. M. Ziman, *Models of Disorder: The Theoretical Physics of Homogeneously Disordered Systems* (Cambridge Univ. Press, Cambridge, 1979; Mir, Moscow, 1982).
3. R. Zallen, *The Physics of Amorphous Solids* (Wiley, New York, 1983).
4. V. A. Polukhin and N. A. Vatolin, *Modeling of Amorphous Metals* (Nauka, Moscow, 1985) [in Russian].
5. V. A. Likhachev and V. E. Shudegov, *Organization Principles of Amorphous Structures* (S.-Peterb. Gos. Univ., St. Petersburg, 1999) [in Russian].
6. H. M. Pak and M. Doyama, J. Fac. Sci., Univ. Tokyo, Ser. B **30**, 111 (1969).
7. A. V. Evteev, A. T. Kosilov, and E. V. Levchenko, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 115 (2002) [JETP Lett. **76**, 104 (2002)].
8. C. J. Smithells, *Metals Reference Book*, 5th ed. (Butterworths, London, 1976; Metallurgiya, Moscow, 1980).
9. R. Yamamoto, H. Matsuoka, and M. Doyama, Phys. Status Solidi A **45**, 305 (1978).
10. D. K. Belashchenko, Fiz. Met. Metalloved. **60**, 1076 (1985).
11. A. V. Evteev and A. T. Kosilov, Rasplavy **1**, 55 (1998).
12. A. V. Evteev and A. T. Kosilov, Rasplavy **4**, 82 (2001).
13. A. V. Evteev, A. T. Kosilov, and E. V. Levchenko, Acta Mater. **51**, 2665 (2003).
14. L. Verlet, Phys. Rev. **159**, 98 (1967).
15. D. K. Belashchenko, Usp. Fiz. Nauk **169**, 361 (1999) [Phys. Usp. **42**, 297 (1999)].
16. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics*, 4th ed. (Nauka, Moscow, 1995; Butterworth, London, 1999), Part 1.
17. A. V. Evteev, A. T. Kosilov, and A. V. Milenin, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 294 (2000) [JETP Lett. **71**, 201 (2000)].
18. A. V. Evteev, A. T. Kosilov, and A. V. Milenin, Fiz. Tverd. Tela (St. Petersburg) **43**, 2187 (2001) [Phys. Solid State **43**, 2284 (2001)].

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 99, No. 3, 2004, pp. 530–538.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 3, 2004, pp. 609–618.

SOLIDS
Structure

Resonant Tunneling of Electrons through Virtual Interference States Formed as a Result of Reflection from the Boundary of Strongly Doped Region of GaAs

Yu. N. Khanin, E. E. Vdovin, and Yu. V. Dubrovskiĭ

Institute of Microelectronic Technology and Ultra-High-Purity Materials, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia

e-mail: khanin@ipmt-hpm.ac.ru

Received March 12, 2004

Abstract—The paper is devoted to analysis of the electron transport through one-barrier GaAs/AlAs/GaAs heterostructures. The oscillating component of transport characteristics of symmetric one-barrier GaAs/AlAs/GaAs heterostructures with spacers, which is associated with resonance tunneling of electrons via virtual states formed in the spacer region of the structures due to reflection of electrons from the n^--GaAs/n^+-GaAs interface and their subsequent interference. It is shown that electrons are predominantly reflected coherently from the boundary of the strongly doped region as in the case of one-dimensional averaged potential of randomly arranged (beginning from this boundary) impurities. It is shown that low-energy virtual resonances are suppressed due to electron scattering as a result of their interaction with longitudinal optical (LO) phonons in the spacer region. *© 2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In this paper, we analyze the vertical transport of electrons in symmetric and asymmetric one-barrier GaAs/AlAs heterostructures with spacers (weakly doped near-barrier layers), associated with resonance tunneling via virtual states formed in the spacer region of the structure as a result of above-the-barrier reflection of electrons from a smooth potential drop at the spacer interface with a strongly doped contact region (n^--GaAs/n^+-GaAs interface) and their subsequent interference with electrons moving towards this interface from the side of the AlAs barrier. In other words, we will consider resonant tunneling via the interference electronic states bounded by the real potential barrier (AlAs layer in the present case), on the one hand, and by the above-the-barrier reflection, on the other, i.e., through the states associated with quantization of longitudinal (coinciding with the growth direction z) motion of electrons in the spacer region (states in a quantum pseudowell). The observation of this effect was briefly reported earlier [1]. In addition, the observation of manifestations of virtual states bounded by the above-the-barrier reflection from sharp (step) and large potential discontinuities (1–10 eV) at heteroboundaries [2, 3] and at the semiconductor– metal interfaces [4, 5] in the transport characteristics of tunnel heterostructures was reported in several experimental works. In the latter of the cited publications, in which electron tunneling in a GaAs/AlAs resonant tunnel diode with an Al Schottky collector were studied and virtual states were formed in the spacer region due to reflection at the boundary between the spacer and the Al collector, the substantial effect of the virtual states on the transport and frequency parameters (above all, the limiting frequency) of a resonant tunneling diode was demonstrated in the temperature range 1.5–300 K. In addition, analysis of the temperature and energy dependences of the number and magnitude of resonant singularities in the transport characteristics, which are associated with resonant tunneling via virtual states (amplitudes of virtual resonances) provided valuable information on the electron distribution in the emitter of a resonant tunneling diode, the mechanisms and rates of scattering in the spacer region, and the reflection properties of the semiconductor–metal (GaAs–Al) interface. Unsuccessful attempts at discovering such resonances undertaken in [6], where resonant tunnel diodes differing from those in [4] only in the method of obtaining the layer of Al collector point towards the strong dependence of the amplitudes of virtual resonances on the properties of reflecting boundary. In the former case [4], the layer was grown directly in the course of molecular beam epitaxy (MBE), while in the latter case [6], the layer was grown in the process following the heterostructure growth. The absence of virtual resonances [6] was attributed to substantial electron scattering by imperfections of the GaAs/Al interface or, according to an alternative opinion, to the fact that the boundary with the Al layer is spatially inhomo-

geneous and fails to ensure the mirror reflection required for the formation of virtual states.

In light of the above arguments, the observation of virtual states associated with the reflection from the apparently relatively weak and broad potential drop across the blurred n^--GaAs/n^+-GaAs interface is rather unexpected and points to the importance of analyzing the influence on the transport of electron reflection from such interfaces, which are inevitably present in a large number of various heterostructures.

The most important of our results pertaining to the above scope of problems form the basis of this publication and can be preliminarily and briefly formulated as follows.

1. We discovered the oscillating component of the transport characteristics of symmetric one-barrier GaAs/AlAs/GaAs heterostructures with spacers, which is associated with resonant tunneling of electrons via virtual states formed in the spacer region of the structures due to reflection of electrons from the n^--GaAs/n^+-GaAs interface. Experimental singularities were identified from a comparison with the calculated values of resonant voltages corresponding to alternate opening of the resonant tunneling channels via the states in a model quantum well and, hence, indicating the position on the transport characteristics of resonant singularities corresponding to these processes.

2. The analysis of the dependence of the virtual resonance amplitude on the bias voltage, which was based on a comparison with the calculated dependence, provided information on the mechanism of reflection from the n^--GaAs/n^+-GaAs interface and on the processes of electron scattering in the spacer region (pseudowells). It was shown that electrons are predominantly reflected from the boundary of the strongly doped region as from a one-dimensional (along the transport direction z) averaged potential of randomly distributed (beginning from this boundary) impurities without being affected by fluctuations and nonuniformities of the potential in the xy plane, which can be expected in actual practice. Analogously to the roughnesses and inhomogeneities of the interfaces in traditional resonant tunnel diodes, such fluctuations are possible and may lead to electron scattering and to blurring (suppression and broadening) of tunnel resonances caused by electron scattering (see [7], pp. 44 and 55). In other words, it was proved that reflection from the boundary of the strongly doped region predominantly occurs coherently. In addition it was shown that the suppression of low-energy virtual resonances is due to scattering of electrons as a result of their interaction with LO phonons in the spacer region.

3. We studied magnetically induced tunneling of electrons in symmetric one-barrier GaAs/AlAs heterostructures with a spacer in a magnetic field perpendicular to the transport direction ($B \perp I$) up to 8 T. Analysis of the evolution of the resonant structure with a magnetic field revealed complete analogy with the situation observed in experiments on electron tunneling via states in wide quantum wells of traditional two-barrier resonant tunnel GaAs/AlAs heterostructures [8] and provided additional confirmation of the correctness of identification of peculiarities in the transport characteristics observed by us as a consequence of the manifestation of resonant tunneling via virtual states in the spacer region.

4. We studied the electron transport in asymmetric GaAs/AlAs heterostructures with a two-step barrier and spacers. It was found that the peculiarities in the transport properties observed for opposite directions of the bias voltage are associated with resonant tunneling via virtual states, which are bounded in the regions of the spacer and the lower barrier step with different widths (via virtual states in pseudowells with different widths). Thus, we experimentally demonstrated the dependence of the "period" of the observed experimental structures (resonance staircase) on the width of the quantum pseudowell; this completely eliminated a certain indeterminacy in the results of identification of peculiarities in the transport characteristics of symmetric heterostructures, which appears due to insufficient knowledge of the parameters of the structures (exact position of the n^-/n^+ interface) as well as physical factors (e.g., the extent of conservation of the wave vector component $k_\parallel = (k_x, k_y)$ perpendicular to the transport direction).

We will now consider each of these results in greater detail, after describing the samples and the measuring technique used in our experiments.

2. SAMPLES AND MEASURING TECHNIQUE

The samples studied here were one-barrier heterodiodes of two types, manufactures on the basis of symmetric GaAs/AlAs/GaAs heterostructures (with different thicknesses of the barrier AlAs layers) and asymmetric GaAs/$Al_{0.4}Ga_{0.6}As$/$Al_{0.03}Ga_{0.97}As$/GaAs heterostructures (with a two-step barrier) grown using molecular-beam epitaxy on highly doped n^+-GaAs substrates with the (100) orientation at 570°C. As a doping impurity, we used silicon with a concentration of 2×10^{18} cm^{-3} in the substrates.

The following description will mainly concern symmetric samples; samples of the second type together with the corresponding experimental results will be described at a later stage. Thus, symmetric heterostructures consisted of the following sequence of layers: 2×10^{18} cm^{-3} n^+-GaAs buffer of thickness 0.4 μm; 2×10^{16} cm^{-3} n^--GaAs spacer of thickness 50 nm; undoped GaAs spacer of thickness 10 nm; undoped AlAs barrier of thickness 2.5, 3.5 or 5.0 nm; undoped GaAs spacer of thickness 10 nm; 2×10^{16} cm^{-3} n^--GaAs spacer of thickness 50 nm; and 2×10^{18} cm^{-3} n^+-GaAs contact layer of thickness 0.4 μm. Ohmic contacts were prepared by consecutive deposition of AuGe/Ni/Au layers and annealing at T = 400°C. To create a mesostructure

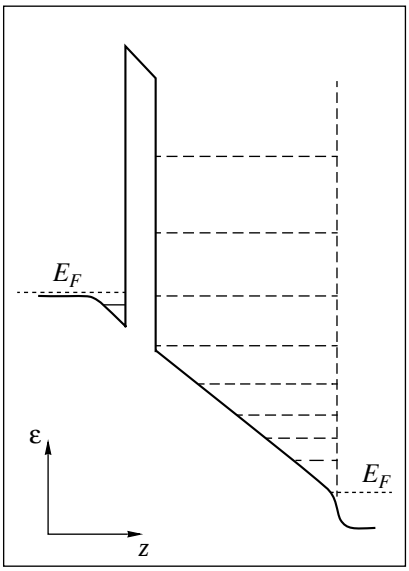

Fig. 1. Schematic band diagram of the experimental one-barrier heterostructure under a positive bias voltage.

100 μm in diameter, we used the standard chemical etching technique.

The schematic profile of the conduction band of a symmetric experimental structure under a bias voltage is shown in Fig. 1. The vertical dashed line in the figure shows the "wall" of the quantum pseudowell, while horizontal dashed line show the states (energy levels) in this pseudowell. The experimental samples were prepared in such a way that the application of an external voltage V_b leads to the formation of the so-called accumulation layer in the vicinity of the AlAs barrier (see Fig. 1); this layer has the form of a pseudo-triangular quantum well with two-dimensional states filled with electrons, which form a two-dimensional electron gas. We confirmed the presence of the accumulation layer in the samples by analyzing magnetoscillations of the Shubnikov–de Haas type [9]. In such a situation, resonant tunneling via two-dimensional states in a quantum pseudowell is associated with tunnel transitions of electrons from two-dimensional states of the accumulation layer via these states. An increase in voltage V_b lowers the energies of the states in the pseudowell relative to the states in the accumulation layer and successively opens resonant tunneling channels at instants of coincidence of the energies of resonances in the corresponding subbands, which is manifested in the form of a sequence of singularities (resonance staircase) in the transport characteristics. It should be noted that, in the case of tunneling with the conservation of the transverse wave vector component k_{xy} considered here, the resonance singularity of the current–voltage characteristic (current resonance) must have Lorentzian or cosinusoidal shape (depending on the broadening of the resonance states) and the corresponding derivatives.

The results described here were based on our measurements of the dependence of the differential conductivity and second derivative of current with respect to voltage across the experimental samples on the applied voltage in the temperature range 4.2–150 K in a magnetic field up to 8 T. These dependences were measured using the standard modulation technique.

2.1. Transport Characteristics of Symmetric GaAs/AlAs/GaAs Samples and Identification of the Observed Resonant Structure

Figure 2 shows the dependences of dI/dV and d^2I/dV^2 on the voltage applied to symmetric GaAs/AlAs/GaAs samples with 3.5-nm-wide Al barriers, which were measured at $T = 4.2$ K in zero magnetic field. These parameters demonstrate the presence of a substantial oscillating component associated (see below) with resonant tunneling of electrons from 2D states of the accumulation layer via 2D virtual states formed in the spacer region of the structures as a result of reflection of electrons from the n^--GaAs/n^+-GaAs interface and their subsequent interference (via 2D states in the quantum pseudowell). Samples with 2.5- and 5-nm-wide AlAs barriers possessed analogous transport properties, the only difference being that resonant oscillations in 5-nm-wide barriers were considerably suppressed under negative bias voltages. It should be noted that the amplitudes of oscillations observed in samples with 3.5-nm-wide barriers on the negative branch of the dI/dV dependence on V are slightly smaller than on the positive branch. This is apparently due to the difference in the sharpnesses of the concentration fronts at the n^+-GaAs/n^--GaAs and the n^--GaAs/n^+-GaAs interfaces (lower and upper reflecting boundaries in the order of layer growth) due to additional blurring of the n^+/n^- interface associated with diffusion during the growth, which is responsible for the corresponding difference in the reflectivities of these interfaces. The decrease in the amplitudes of resonance oscillations with increasing bias voltage is primarily associated with a decrease in the lifetime (an increase in the energy broadening) of the corresponding virtual states due to a decrease in the coefficient of the above-the-barrier reflection at the n^-/n^+ interface with increasing energy.

The identification of experimental peculiarities as a manifestation of resonant tunnelling via virtual states in the spacer region (virtual resonances) was carried out by comparing the experimental characteristics with the calculated resonant voltages, indicating the expected position of resonance singularities of a certain type on the transport characteristics, which correspond to coincidences of energies (resonances) of subbands of the accumulation layer and (actual) quantum well in the model tunnel structures. The model structure differed from the experimental one only in the presence of an additional thin barrier at the n^+/n^- interface. The pro-

files of the bottom of the conduction band, the energy levels corresponding to the states in the quantum well, and the characteristics (energy level and the electron concentration specified by it) of the accumulation layer as functions of the applied voltage were determined from the self-consistent solution of the Poisson and Schrödinger equations. In this case, the correctness of the calculation of accumulation layer parameters was verified by comparison with the experimental voltage dependence of the electron concentration in the accumulation layer, which was derived from an analysis of the Shubnikov–de Haas oscillations in a magnetic field parallel to the current ($B \parallel I$) [9].

It should be noted that the above condition of the coincidence of the energies of the subbands in the accumulation layer and the quantum well is convenient for describing resonant tunneling with conservation of the transverse (relative to the current) wave vector component k_{xy} and corresponds to maxima on the current–voltage (I–V) characteristic (minima on the d^2I/dV^2–V characteristics). The results of calculating resonance voltages of the model structure with a 60.5-nm-wide quantum well are shown by triangles in Fig. 2b. The calculated voltages corresponded to tunneling via eight states in the quantum well, starting from the third one. Tunneling via two lower states is not observed due to the presence of electron-enriched layers in our structures on both sides of the barriers for zero bias voltage (it is the tunneling between these layers that determines the transport characteristics for low bias voltages). This was described in detail in our previous publication [10]. The identification carried out in this way is not rigorous in view of the dependence on technological parameters used in simulation (e.g., the position of the reflecting n^-/n^+ interface) and a tunneling parameter such as the degree of conservation of k_{xy}, which are not known exactly. A proof of the correctness of the identification was obtained with the help of additional study of tunneling in a magnetic field in symmetric structures and tunneling in asymmetric structures with a two-step barrier; these structures will be described in Sections 5 and 6 of this article.

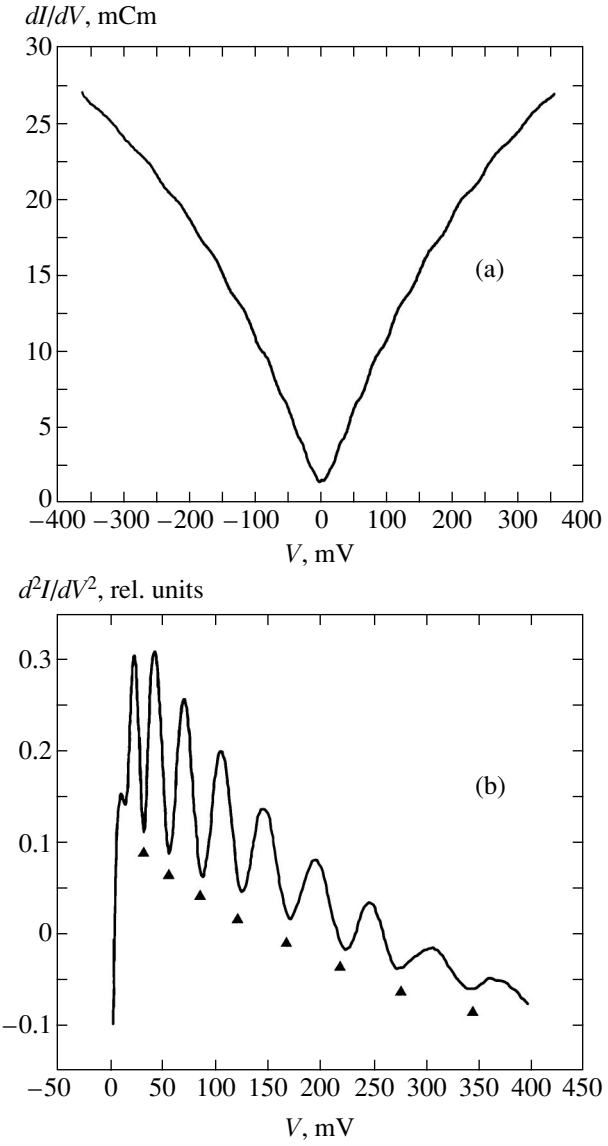

Fig. 2. (a) Experimental dependence of dI/dV on V for a symmetric one-barrier heterostructure with a 3.5-nm-thick barrier, measured at $T = 4.2$ K in zero magnetic field and (b) the positive branch of the d^2I/dV^2 dependence on V. Triangles indicate the calculated values of threshold voltages.

3. ANALYSIS OF AMPLITUDES OF VIRTUAL RESONANCES AND MECHANISMS OF ELECTRON REFLECTION AND SCATTERING

We determined the predominant mechanism of reflection of electrons from the n^--GaAs/n^+-GaAs interface and the mechanism of electron scattering leading to the suppression of low-energy virtual resonances from comparison of relative amplitudes ($\Delta I/I$) of experimentally observed resonance singularities with the values calculated in the following manner.

It follows from the general theory of resonant tunneling that the relative amplitude of the current resonance (resonant singularity on the I–V curve) in the case of a quantum well bounded by barriers with strongly differing transparency (this also applies to our experimental situation with a pseudowell [11]) is determined by the reflection coefficient R of the most transparent barrier [11–14]:

$$\frac{\Delta I}{I} = R^{1/2}. \qquad (1)$$

In the present case, the role of this barrier is played by the above-the-barrier reflection at the n^-/n^+ interface. The reflecting contact layer n^+ was simulated in the form of the half-space $z \geq W$, where W is the length of the spacer region, with randomly distributed impurities

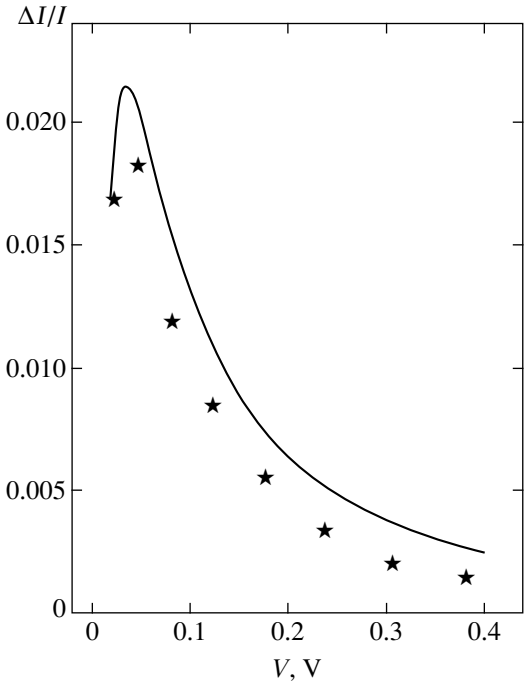

Fig. 3. Experimental (stars) and calculated (solid curve) values of the relative amplitudes of the first eight virtual resonances. The details of calculation are described in text.

with an average volume concentration of $N_D = 2 \times 10^{18}$ cm^{-3}. We assume that the other half-space ($z < W$) is free of impurities. The amplitude of the wave $R^{1/2}$ reflected from such a layer is calculated here in the framework of the Born approximation [15],

$$R^{1/2} = \left| \frac{im^*}{\hbar^2 k_z} \int_W^\infty u(z) e^{-2ikz} dz \right| \quad (2)$$

(m^* and k are the effective mass and the wave vector of an electron arriving at the reflecting boundary and $\hbar$ is the Planck constant), using as the scattering potential the one-dimensional averaged potential of the half-space with randomly distributed impurities,

$$u(z) = \frac{2\pi e^2 N_D}{\varepsilon} (Z - W)^2, \quad (3)$$

where ε is the permittivity. As a result of substitutions and integration, we obtained the following expression for the relative amplitude of the current resonance:

$$\frac{\Delta I}{I} = \frac{4\pi m^* e^2 N_D}{\hbar^2 \varepsilon k^4}. \quad (4)$$

The choice of candidates to play the role of mechanisms of electron scattering in the spacer region was made from comparison of electron lifetimes in virtual states with characteristic times of various scattering processes. In view of the short lifetime of electrons in virtual states ($\tau \le 0.2$ ps), the only real candidate was electron scattering as a result of interaction with LO phonons in GaAs with the characteristic time $\tau_{LO} \approx 0.18$ ps [5, 7, 16, 17]. A similar procedure was carried out in other experimental works, in which tunneling via short-lived states in quantum wells of GaAs was studied [5, 16]. We took into account the scattering from optical phonons by introducing factor $\exp(-\tau/\tau_{LO})$ into Eq. (4) analogously to [16]. We assumed that the lifetime τ was equal to the time of electron passage through the spacer region; according to [5], this is valid for states bounded on one side by a barrier with a small reflectance $R \ll 1$. Thus, allowance for the electron–phonon interaction led to the expression

$$\frac{\Delta I}{I} = \frac{4\pi m^* e^2 N_D}{\varepsilon \hbar^2 k^4} \exp\left(-\frac{\tau}{\tau_{LO}}\right), \quad (5)$$

which was used for calculating the voltage dependence of relative amplitudes.

Before comparing the theory and experiments, it should be noted that, as a result of application of one-dimensional averaged potential (3) in calculations, final expression (5) outside the narrow (see below) energy range of the exponential factor (in which the electron–phonon scattering is taken into account), describes the upper boundary of the relative amplitude of virtual resonances. This is due to the fact that we disregard possible fluctuations of reflecting potential in the *xy* plane and the presence of a transient layer with a variable impurity concentration, which may lead to electron scattering and resonance blurring associated with it. Thus, considering that $\Delta I/I$ is a measure of participation of scattering in resonance tunneling processes [5, 7, 18], we can conclude that a comparison of the dependence of $\Delta I/I$ on V constructed in accordance with relation (5) with experimental values of relative amplitudes of resonances will make it possible to determine the extent of integrated effect of electron scattering in the region of the reflecting boundary on the process of reflection from the real n^-/n^+ interface. It should be remarked for clarity that although $\Delta I/I$ is not an exact quantitative measure of the effect of scattering on the tunnel transport (measure of coherence), it was demonstrated theoretically (using, for example, the Breit–Wigner formalism) that the attainment of the equality of broadenings Γ_s and Γ_i associated with scattering and lifetime of the state corresponding to the boundary between the coherent and predominantly incoherent transport is accompanied with more than twofold reduction of the value of $\Delta I/I$ (see [7], p. 46).

The calculated dependence of $\Delta I/I$ on V and averaged values of relative amplitudes are shown in Fig. 3 for seven experimental samples. The deviation from the experimental values does not exceed 5%. It can be seen from the figure that the suppression of the amplitudes of

the first two experimental resonances (predominantly the first resonance corresponding to $\tau \le \tau_{LO}$) is satisfactorily described by the theoretical curve. The behavior of this curve in the given energy range (where $\tau \le \tau_{LO}$) is determined to a considerable extent by the exponential factor, allowing for the effect of the electron–phonon scattering. Such a suppression of the low-energy resonances as a result of the electron–phonon scattering was observed earlier when tunneling in a GaAs/AlAs resonant-tunnel diode with broad quantum wells was studied [16]. In this work, the increase in the amplitudes of "suppressed" resonances in a field parallel to current ($B \parallel I$) was reported. This effect was explained (on the basis of numerical calculations) by a decrease in the rate of scattering $\hbar/\tau_B$ as a result of limitation imposed on the number of final states to which electrons can be scattered. We also observed an increase in the amplitudes of the first two experimental resonances (predominantly the first) in field $B \parallel I$ (the corresponding figure is not given here), which serves as an additional proof of the assumption that the electron–phonon scattering is responsible for the suppression of low-energy resonances.

It can also be seen from Fig. 3 that the difference between the relative amplitudes of the next four resonances and the calculated values does not exceed 25% and increases with voltage. These resonances lie in the energy interval where $\tau \ll \tau_{LO}$ and the exponential factor, taking into account scattering, does not noticeably affect the dependence of $\Delta I/I$ on V. Consequently, taking into account the above arguments about the approaches to comparison of theory with experiment, we can conclude that the observed scale of the difference between the theoretical and experimental amplitudes of resonances indicates that the reflection of electrons from the boundary n^+ of the strongly doped region is predominantly nondissipative (coherent). To be more specific, we conclude that electrons are reflected from this boundary predominantly as from the averaged one-dimensional potential of the half-space with randomly distributed impurities, without experiencing strong scattering at inhomogeneities of the reflecting potential in the xy plane and in the transient layer having a variable concentration of donors.

Finally, it should be noted that a stronger (~50%) difference from the calculated values of the amplitudes of two high-energy resonances can be due to strong suppression of experimental resonances because of stronger involvement of scattering from the above-mentioned fluctuations of the reflecting potential and in the transient layer during reflection upon a decrease in the electron wavelength. Another reason for the discrepancy is incorrect application of the averaged potential of impurities in the given energy range, when the electron wavelength becomes smaller than the mean distance between impurities ($\lambda < N_D^{-1/3}$), which leads to exaggerated values of the amplitudes. The removal of this indeterminacy requires that calculations be made at least by using more complicated model potentials taking into account the microscopic structure of the reflecting boundary. However, this is beyond the scope of the problems considered here and is not reflected in our conclusions.

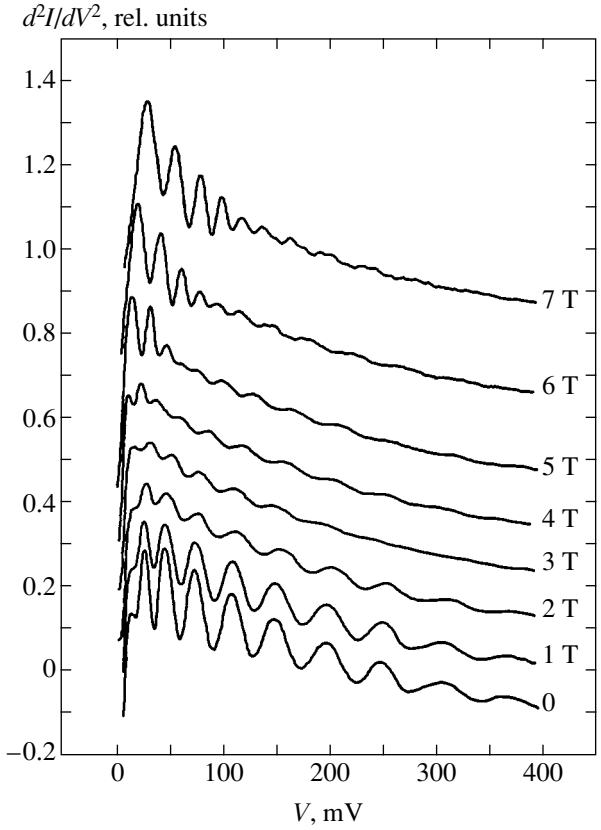

Fig. 4. Experimental dependences of d^2I/dV^2 on V for a one-barrier structure with a 3.5-nm barrier in a transverse magnetic field in the range from 0 to 7 T.

4. MAGNETOTUNNELING IN SYMMETRIC GaAs/AlAs/GaAs SAMPLES IN A MAGNETIC FIELD TRANSVERSE TO CURRENT

Figure 4 shows a family of dependences of d^2I/dV^2 on V for a symmetric sample with a 3.5-nm-wide barrier, which were measured at 4.2 K in a magnetic field transverse to the current ($B \perp I$) up to 7 T. It can be seen that the increase in the magnetic field first leads to a displacement of the oscillatory structure towards higher voltages and to its attenuation down to complete suppression at $B \approx 2.5$ T. Such a behavior of the resonance structure can be explained by the change in the resonance conditions for tunneling between 2D states with conservation of k_{xy} in the magnetic field [19]. The change in the resonance conditions $\Delta E = (eBd)^2/2m^*$ is a consequence of the change in the transverse component of the wave vector of a tunneling electron under the effect of the Lorentz force by $\Delta k = eBd/\hbar$, where

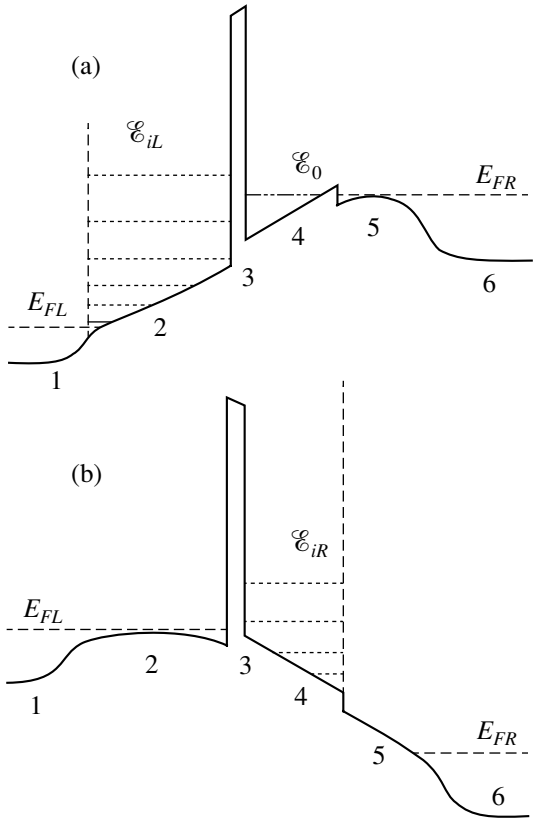

Fig. 5. Schematic band diagram of a heterostructure with a step barrier (a) for a negative and (b) for a positive bias voltage; E_F are the Fermi levels in 3D contacts; $\mathscr{E}_0$ is the size quantization level in the triangular well induced by an electric field; $\mathscr{E}_{iR}$ and $\mathscr{E}_{iL}$ are size quantization levels in pseudowells. Figures 1–6 denote various layers of the structure, described in text.

d is the tunneling length; in the first approximation, we assume that d is equal to the distance between the centers of the quantum wells whose 2D states participate in the tunneling process under investigation. The velocity $\Delta V/\Delta B$ of the resonance structure in a magnetic field calculated in accordance with the above arguments amounts to 6 mV/T under the assumption that $d \approx$ 40 nm and is in good agreement with the experimental value. When the magnetic field attains a value of $B \approx$ 3 T, the d^2I/dV^2 vs. V characteristic acquires a new oscillatory structure in which two components can subsequently be singled out at B = 7 T, viz., the strong "low-energy" component (up to $V \approx 120$ mV) and the weak "high-energy" component (behind $V \approx 120$ mV) oscillatory structures. A similar evolution of the resonance structure with $B \perp I$, including the low-field displacement and suppression of the initial resonances (at $B = 0$) associated with a change in the resonance conditions for tunneling between 2D states due to the Lorentz force and the formation of a new resonance structure upon a further increase in the magnetic field was observed earlier while studying tunneling in two-barrier structures with broad quantum wells [8]. The emergence of the new resonance structure was associated with resonant tunneling via hybrid magnetoelectric states in a quantum well; a semiclassical analog of one type of such states (corresponding to low-energy resonances) are electron orbits hopping along the intrawell boundary of the emitter barrier. However, we will not consider here the classification of magnetoelectric resonances.

Thus, the modification of the oscillatory structure of (d^2I/dV^2–V) characteristics with $B \perp I$ considered here and its analogy with that observed in magnetotunneling via 2D states of broad quantum wells confirm the validity of using the resonance conditions for tunneling between states of 2D subbands with conservation of k_{xy} for identifying singularities in the oscillatory structure at $B = 0$, as well as the results of identifying such singularities as a manifestation of resonant tunneling via virtual 2D states in a quantum pseudowell.

5. TUNNEL RESONANCES IN ASYMMETRIC STRUCTURES WITH A TWO-STEP BARRIER

In this section, we present the results of experimental investigation of electron transport through the asymmetric GaAs/$Al_{0.4}Ga_{0.6}As$/$Al_{0.03}Ga_{0.97}As$/GaAs heterostructure with a two-step barrier and spacers (see Fig. 5). As a result of experiments, singularities in the transport characteristics associated with resonant tunneling via virtual states ($\mathscr{E}_{iL}$ and $\mathscr{E}_{iR}$ in Fig. 5) in two quantum pseudowells of different thickness were detected. One of the pseudowells was bounded by the upper step of the barrier and the reflection from the n^+-GaAs/n^--GaAs interface, while the other was bounded by the upper step of the barrier and the reflection from the $Al_{0.03}Ga_{0.97}As/n^-$-GaAs heteroboundary. In addition, resonant tunneling via the states of the lower subband in the induced triangular quantum well was detected (the bottom of the subband is denoted by $\mathscr{E}_0$ in Fig. 5); this tunneling will not be described here in detail. As in the case of symmetric samples, tunneling via resonance states occurred from two-dimensional states of the accumulation layers formed with opposite voltages at the opposite sides of the two-step barrier. Note that the $Al_{0.4}Ga_{0.6}As$ barrier layer used in the given structures made it possible to eliminate the effect of the X valley on the electron transport.

The samples used in these experiments were prepared on the basis of heterostructures grown by the molecular beam epitaxy method on a highly doped n^+-GaAs substrate and having the following sequence of the layers: 2×10^{18} cm^{-3} n^+-GaAs buffer layer of thickness 0.5 μm (region 1 in Fig. 5); 2×10^{16} cm^{-3} n^--GaAs spacer of thickness 50 nm; undoped GaAs spacer of thickness 10 nm (the two latter layers form region 2 in Fig. 5); undoped barrier layer of

$Al_{0.4}Ga_{0.6}As$ of thickness 5 nm (region 3 in Fig. 5); undoped barrier layer of $Al_{0.03}Ga_{0.97}As$ of thickness 30 nm (region 4 in Fig. 5); 2×10^{16} cm^{-3} n^--GaAs spacer of thickness 30 nm (region 5 in Fig. 5); and 2×10^{18} cm^{-3} n^+-GaAs contact layer of thickness 0.2 μm (region 6 in Fig. 5). Silicon was used as the doping impurity. The sample preparation technique was the same as that described in Section 2.

Figure 6 shows the oscillating component of the experimental (d^2V/dI^2–V) characteristic of an asymmetric sample in zero magnetic field at T = 4.2 K. Large-scale components of the characteristic, which are due to considerable zero-point anomaly and resonance singularity associated with tunneling via states of the two-dimensional subband in the triangular quantum well (disregarded here) are subtracted for better visualization when oscillatory structures for opposite voltages are compared. The observed oscillatory structures are due to resonant tunneling of electrons via virtual states in quantum pseudowells of different thicknesses (60 and 30 nm; see Fig. 5). This immediately follows from the substantial difference in the "frequencies" of experimental oscillations (heights of the steps in staircases of resonances) on the negative and positive branches of the characteristics. The predicted values of resonance voltages calculated in the same way as in Section 3 are shown in Fig. 6 by triangles and demonstrate satisfactory agreement with experiment.

The absence of resonances corresponding to interference states emerging due to reflection from the n^-/n^+ interface in the region of negative bias voltages is worth noting. This can be associated with a decrease in the probability of constructive interference of electrons due to the presence of additional scattering interfaces on their path ($Al_{0.03}Ga_{0.97}As/n^-$-GaAs in the present case).

For both bias voltages, the values of relative amplitudes $\Delta I/I$ of the resonances do not exceed 1%, which is slightly smaller than the value observed for symmetric structures. Analysis of the dependences of $\Delta I/I$ on V for asymmetric structures was complicated to a considerable extent by the presence of the lower step of the barrier, which led to a strong voltage dependence of the integral barrier transparency (a description of this dependence involves considerable difficulties) and was not fruitful.

Concluding the section, we note that the behavior of the oscillatory structure in a magnetic field $B \perp I$ was similar to that observed for symmetric samples, which additionally confirmed the assumption about the two-dimensional nature of the states via which the tunneling takes place.

Thus, the analysis of the electron transport in asymmetric structures with a two-step barrier revealed oscillatory components of the transport characteristics with strongly differing "frequencies." These frequencies are associated with resonant tunneling of electrons through quantum pseudowells of various width, thus demonstrating the dependence of the heights of the steps in the staircase of virtual states on the width of the quantum pseudowell bounding these states. This completely eliminated a certain ambiguity in the results of identification of the resonance structure of transport characteristics for symmetric samples.

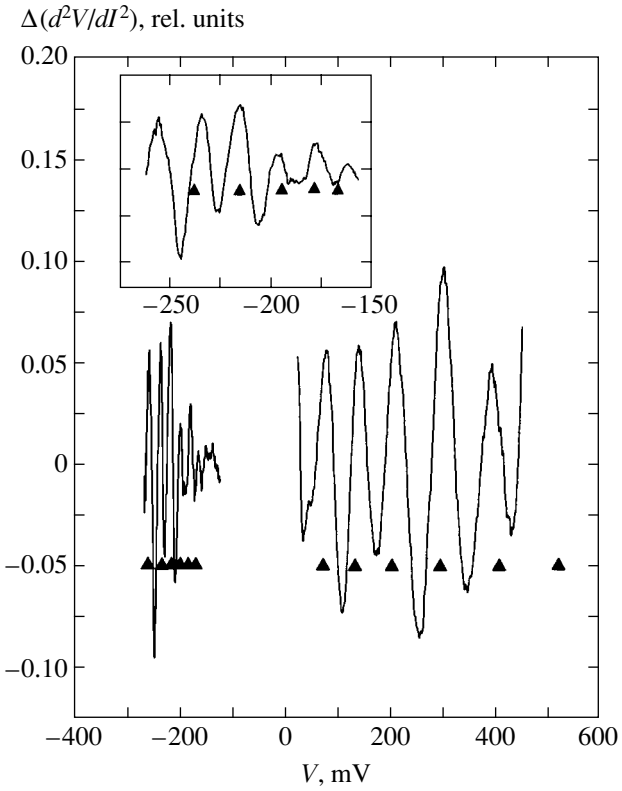

Fig. 6. Dependence of the second derivative of voltage with respect to current, d^2V/dI^2, with subtracted monotonic component, on the bias voltage V. Triangles show the results of calculation of the positions of tunneling thresholds. The inset shows the resonances under a negative bias voltage.

6. CONCLUSIONS

Thus, as a result of investigations described in this paper, we observed predominantly coherent reflection of electrons from the n^--GaAs/n^+-GaAs interface, leading to the formation of interference electron states in weakly doped spacer regions of heterostructures. Resonant tunneling via such states was manifested in the form of the oscillating component of transport characteristics up to $T \approx 150$ K.

REFERENCES

1. E. E. Vdovin, Yu. V. Dubrovskiĭ, I. A. Larkin, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **61**, 566 (1995) [JETP Lett. **61**, 576 (1995)].
2. T. W. Hickmott, Phys. Rev. B **40**, 11683 (1989).
3. R. C. Potter and A. A. Lakhani, Appl. Phys. Lett. **52**, 1349 (1988).

4. M. V. Weckworth, J. P. A. van der Wagt, and J. S. Harris, J. Vac. Sci. Technol. B **12**, 1303 (1994).
5. A. J. North, E. H. Linfield, M. Y. Simmons, *et al.*, Phys. Rev. B **57**, 1847 (1998).
6. Y. Konishi, S. T. Allen, M. Reddy, and M. J. W. Rodwell, Solid-State Electron. **36**, 1673 (1993).
7. H. Mizuta and T. Tanoue, *The Physics and Applications of Resonant Tunneling Diodes* (Cambridge Univ. Press, New York, 1995).
8. M. Helm, F. M. Peeters, P. England, *et al.*, Phys. Rev. B **39**, 3427 (1989).
9. Yu. V. Dubrovskii, Yu. N. Khanin, I. A. Larkin, *et al.*, Phys. Rev. B **50**, 4897 (1994).
10. Yu. N. Khanin, Yu. V. Dubrovskiĭ, and E. E. Vdovin, Fiz. Tekh. Poluprovodn. (St. Petersburg) **37**, 717 (2003) [Semiconductors **37**, 692 (2003)].
11. V. I. Fal'ko and S. V. Meshkov, Semicond. Sci. Technol. **6**, 196 (1991).
12. J. H. Davies, *The Physics of Low-Dimensional Semiconductors: An Introduction* (Cambridge Univ. Press, Cambridge, 1998).
13. M. Buttiker, IBM J. Res. Dev. **32**, 63 (1988).
14. B. Ricco and M. Ya. Azbel, Phys. Rev. B **29**, 1970 (1984).
15. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory*, 2nd ed. (Fizmatgiz, Moscow, 1963; Pergamon, New York, 1977).
16. A. F. J. Levi, R. J. Span, and J. H. English, Phys. Rev. B **36**, 9402 (1987).
17. R. Ferreira and G. Bastard, Phys. Rev. B **40**, 1074 (1989).
18. M. Buttiker, in *Resonant Tunnelling in Semiconductors*, Ed. by L. L. Chang *et al.* (Plenum, New York, 1991).
19. J. M. Smith, P. C. Klipstein, R. Grey, and G. Hill, Phys. Rev. B **58**, 4708 (1998).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 99, No. 3, 2004, pp. 539–543.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 3, 2004, pp. 619–624.

SOLIDS
Structure

Universality in the Partially Anisotropic Three-Dimensional Ising Lattice¶

M. A. Yurishchev
Institute of Problems of Chemical Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia
e-mail: yur@itp.ac.ru, yur@icp.ac.ru
Received March 16, 2004

Abstract—Using transfer-matrix extended phenomenological renormalization-group methods, we study the critical properties of the spin-1/2 Ising model on a simple-cubic lattice with partly anisotropic coupling strengths $\vec{J} = (J', J', J)$. The universality of both fundamental critical exponents y_t and y_h is confirmed. It is shown that the critical finite-size scaling amplitude ratios $U = A_{\chi^{(4)}}A_\kappa/A_\chi^2$, $Y_1 = A_{\kappa''}/A_\chi$, and $Y_2 = A_{\kappa^{(4)}}/A_{\chi^{(4)}}$ are independent of the lattice anisotropy parameter $\Delta = J'/J$. For the Y_2 invariant of the three-dimensional Ising universality class, we give the first quantitative estimate $Y_2 \approx 2.013$ (shape $L \times L \times \infty$, periodic boundary conditions in both transverse directions). © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The phenomenological renormalization-group (RG) method in which the transfer-matrix technique and finite-size scaling (FSS) ideas are combined is a powerful tool for investigation of critical properties in different two-dimensional systems [1, 2]. Unfortunately, its application in three or more dimensions is sharply retarded due to huge sizes of the transfer matrices arising in approximations of d-dimensional lattices by $L^{d-1} \times \infty$ subsystems.

Indeed, even in the simplest case of systems with only two states of a site (the spin-1/2 Ising model), the size of the transfer matrix in three dimensions ($d = 3$) increases as 2^{L^2} (instead of the significantly more sparing law 2^L in two dimensions). Hence, for the $3 \times 3 \times \infty$ cluster, the eigenproblem of the 512×512 transfer matrix must be solved; for the $4 \times 4 \times \infty$ subsystem, the problem is for the 65536×65536 matrix; and for the $5 \times 5 \times \infty$ cluster, it is required to find the eigenvalues and eigenvectors of dense matrices with huge sizes of 33554432 by 33554432.

One can solve the full eigenproblem for the transfer matrices of Ising parallelepipeds $L \times L \times \infty$ with the side length $L \le 4$. Our aim in this paper is to use such solutions with the maximum effect and extract as much accurate information as possible about the physical properties of the bulk system.

The ordinary phenomenological RG is based on the FSS equations for correlation lengths [1, 2]. However, it is known [3–5] that the phenomenological RG can be built up using other quantities with a power divergence at the phase transition point. It is remarkable that such modified renormalizations can provide more precise results with the same sizes of subsystems [6].

In this paper, we calculate the values of different invariants of the 3D Ising universality class and discuss their universal and extrauniversal properties.

2. BASIC EQUATIONS

We start from the ordinary FSS equations [1, 2] for the inverse correlation length $\kappa_L(t, h)$ and the singular part of the dimensionless free-energy density $f_L^s(t, h)$, but we write them for the derivatives with respect to a reduced temperature $t = (T - T_c)/T_c$ and external field h,

$$\kappa_L^{(m,n)}(t, h) = b^{my_t + ny_h - 1}\kappa_{L/b}^{(m,n)}(t', h') \quad (1)$$

and

$$f_L^{s(m,n)}(t, h) = b^{my_t + ny_h - d}f_{L/b}^{s(m,n)}(t', h'). \quad (2)$$

Here,

$$\kappa_L^{(m,n)}(t, h) = \frac{\partial^{m+n}\kappa_L}{\partial t^m \partial h^n}$$

and similarly for $f_L^{s(m,n)}$; y_t and y_h are the thermal and magnetic critical exponents of the system, respectively; and $b = L/L'$ is the rescaling factor. In deriving Eqs. (1) and (2), we used a linearized form of the RG equations $t' \approx b^{y_t}t$ and $h' \approx b^{y_h}h$.

¶ This article was submitted by the author in English.

In the traditional phenomenological RG theory [1, 2], Eq. (1) with $m = n = 0$ is considered as an RG mapping $(t, h) \longrightarrow (t', h')$ for a cluster pair (L, L'). The critical temperature T_c is then estimated from the equation

$$L\kappa_L(T_c) = L'\kappa_{L'}(T_c). \quad (3)$$

The phenomenological renormalization $(t, h) \longrightarrow (t', h')$ can also be realized by using any of relations (1) and (2) or a combination of them. The author has shown [6] that some of such extended renormalizations lead to more rapid convergence in L than the standard phenomenological RG transformation. In particular, test examples on fully isotropic systems [6] have shown that the relations

$$\left.\frac{\kappa_L''}{L^{d-1}\chi_L}\right|_{T_c} = \left.\frac{\kappa_{L'}''}{(L')^{d-1}\chi_{L'}}\right|_{T_c}, \quad (4)$$

$$\left.\frac{\chi_L^{(4)}}{L^d\chi_L^2}\right|_{T_c} = \left.\frac{\chi_{L'}^{(4)}}{(L')^d\chi_{L'}^2}\right|_{T_c} \quad (5)$$

locate T_c more accurately in comparison with the ordinary RG equation (3). In relations (4) and (5), the derivative $\kappa_L'' = \partial^2\kappa_L/\partial h^2$, the zero-field susceptibility $\chi_L = f_L^{s(0,2)}$ and the nonlinear susceptibility $\chi_L^{(4)} = f_L^{s(0,4)}$ can be evaluated by standard formulas via the eigenvalues and eigenvectors of transfer matrices (see, e.g., [7–9]).

To find the thermal critical exponent y_t, we applied two approaches. First, we again used the standard finite-size expression

$$y_t = \frac{\ln[L\dot\kappa_L/(L'\dot\kappa_{L'})]}{\ln(L/L')}, \quad (6)$$

which follows from Eq. (1) with $m = 1$, $n = 0$; $\dot\kappa_L = \partial\kappa_L/\partial t$. Second, we took the formula

$$y_t = \frac{\kappa_{L'}\dot\kappa_L - \kappa_L\dot\kappa_{L'}}{(\kappa_L\kappa_{L'}\dot\kappa_L\dot\kappa_{L'})^{1/2}\ln(L/L')}. \quad (7)$$

This expression is a direct consequence of the well-known Roomany–Wyld approximant to the Callan–Symanzik β-function [2].

To calculate the magnetic critical exponent y_h, we also used two methods,

$$y_h = \frac{d}{2} + \frac{\ln(\chi_L/\chi_{L'})}{2\ln(L/L')} \quad (8)$$

and

$$y_h = \frac{1}{2} + \frac{\ln(\kappa_L''/\kappa_{L'}'')}{2\ln(L/L')} \quad (9)$$

(these finite-size relations follow from Eqs. (1) and (2)).

In addition, we calculated the universal ratios of the critical FSS amplitudes. Ratios of this type can be identified from the Privman–Fisher functional expressions [10]. For the anisotropic systems discussed, they are given by [8]

$$\kappa_L(t, h) = L^{-1}G_0\mathcal{H}(C_1tL^{y_t}, C_2hL^{y_h}), \quad (10)$$

$$f_L^s(t, h) = L^{-d}G_0\mathcal{F}(C_1tL^{y_t}, C_2hL^{y_h}). \quad (11)$$

The scaling functions $\mathcal{H}(x_1, x_2)$ and $\mathcal{F}(x_1, x_2)$ are the same within the limits of a given universality class, but they may depend on the boundary conditions and the shape of the subsystem (a cube, infinitely long parallelepipeds, etc.). Thus, all nonuniversality, including the lattice anisotropy parameter Δ, is absorbed in the geometry prefactor G_0 and metric coefficients C_1 and C_2. The critical amplitude ratios from which the parameters G_0, C_1, and C_2 drop out should be extrauniversal. In particular, the amplitude combinations

$$U = \frac{A_{\chi^{(4)}}A_\kappa}{A_\chi^2} = \frac{\kappa_L\chi_L^{(4)}}{L^{d-1}\chi_L^2} \quad (12)$$

(a Binder-like ratio for the spatially anisotropic systems),

$$Y_1 = \frac{A_{\kappa''}}{A_\chi} = \frac{\kappa_L''}{L^{d-1}\chi_L}, \quad (13)$$

$$Y_2 = \frac{A_{\kappa^{(4)}}}{A_{\chi^{(4)}}} = \frac{\kappa_L^{(4)}}{L^{d-1}\chi_L^{(4)}} \quad (14)$$

are expected to be independent of the lattice anisotropy parameter $\Delta = J'/J$.

3. RESULTS AND DISCUSSION

We have carried out calculations for the subsystems $L \times L \times \infty$ with $L = 3$ and 4. To avoid undesirable surface effects, the periodic boundary conditions were imposed in both transverse directions of parallelepipeds $L \times L \times \infty$. Thus, the transfer matrices for which the eigenproblem was solved were dense matrices of sizes up to 65536×65536. To solve the eigenproblem, we took the internal and lattice symmetries of subsystems into account and used the block-diagonalization method (see, e.g., [7, 9]). Calculations were performed on an 800 MHz Pentium III PC running the FreeBSD operating system.

3.1. Critical Temperature

The critical temperature estimates coming from solutions of transcendental equations (4) and (5) are shown in Table 1.

In the purely isotropic case ($J' = J$), there are high-precision numerical estimates for the critical point of

the three-dimensional Ising model. The most precise values for it have been obtained by Monte Carlo simulations [11, 12]: $K_c = 0.22165459(10)$, i.e., $k_BT_c/J = 1/K_c = 4.5115240(21)$.

One can see from Table 1 that the estimates for $J' = J$ that follow from Eqs. (4) and (5) are the lower and upper bounds, respectively. Therefore, their mean value has an accuracy of 0.01%. We also note that our mean estimate is better than the value $k_BT_c/J = 4.53371$ obtained in [13] (see also [14]) for the fully isotropic lattice using the ordinary phenomenological renormalization of the bars with $L = 4, 5$.

We now discuss the anisotropic case. Here, there is the well-known exact asymptotic formula for the critical temperature [15],

$$\left(\frac{k_BT_c}{J}\right)_{\text{asym}} = \frac{2}{\ln(J/2J') - \ln\ln(J/2J') + O(1)} \quad (15)$$

as $J'/J \longrightarrow 0$. This is a direct consequence of the molecular-field approximation, in which the linear Ising chain is taken as a cluster.

Unfortunately, simple formula (15) yields considerable errors in the region $10^{-3} \le J'/J \le 1$. Its modifications in the spirit of [16],

$$\frac{k_BT_c}{J} \approx \frac{2}{\ln(J/J') - \ln\ln(J/J')},$$

lead to a loss of monotonic convergence as J'/J varies from unity to zero.

We choose infinitely long clusters $L \times L \times \infty$ stretched in the lattice direction with the dominant interaction J. Such a cluster geometry reflects the physical situation in the system. We may therefore expect more precise results for the critical temperature as the anisotropy of the quasi-one-dimensional lattice increases. We may also expect monotonic convergence for the estimates in Eqs. (4) and (5) because there must be physical reasons (finite length of clusters in the longitudinal direction, etc.) for nonmonotonic or oscillatory behavior; they are absent in our approximations. That is, if Eq. (4) yields the lower bound in the most unfavorable case $J' = J$, then it should preserve such behavior for all $J' < J$. Similar arguments are valid for the estimates following from Eq. (5); these are upper estimates.

We note that the mean values in Table 1 are not only better than the estimates of k_BT_c/J calculated with the (3, 4) cluster pair by the standard phenomenological RG method, but also better than their improvements found by means of three-point extrapolations from the sizes $L = 2$, 3, and 4 to the bulk limit [17].

In the range $10^{-2} \le J'/J \le 1$, there are also data for the critical temperature of a simple-cubic Ising lattice that were extracted from the Padé-approximant analysis of the high-temperature series [18]. For $J' = J$, according to these data, $k_BT_c/J = 4.5106$, which is lower by 0.014% in comparison with the results in [12]. For $J'/J = 0.1$, the authors of [18] found the value $k_BT_c/J = 1.343$. This quantity somewhat overestimates the mean value in Table 1. Finally, for $J'/J = 0.01$, the series method [18] yields $k_BT_c/J = 0.65$, which exceeds our lower bound. This is not surprising because the calculations based on the high-temperature series rapidly deteriorate owing to the very limited number (≤11) of terms available in such series for the anisotropic lattices.

Table 1. Lower and upper bounds on the critical temperature and their mean values (improved estimates of k_BT_c/J) in the three-dimensional simple-cubic spin-1/2 Ising lattice vs. $\Delta = J'/J$. Calculations with a cluster pair (3, 4)

Δ	Eq. (4)	Eq. (5)	mean
1.0	4.47965814	4.54424309	4.51195062
0.5	2.91008665	2.94295713	2.92652189
0.1	1.33649605	1.34570054	1.34109829
0.05	1.03544938	1.04144927	1.03844933
0.01	0.65054054	0.65323146	0.65188600
0.005	0.55440490	0.55643112	0.55541801
0.001	0.40743000	0.40859011	0.40801006

Therefore, we may treat the values found from Eqs. (4) and (5) as lower and upper bounds on the real critical temperature. Their mean value for each J'/J yields the best estimate achieved by us in this paper for the reduced critical temperature k_BT_c/J (the last column in Table 1). Hence, its absolute error is not larger in any case than half the difference of the corresponding upper and lower bounds. Using the data in Table 1, we establish that the relative errors for k_BT_c/J monotonically decrease from 0.72% to 0.14% as J'/J goes from 1 to 10^{-3}.

3.2. Invariants of the 3D Ising Universality Class

With the improved estimates for the critical temperature of anisotropic simple-cubic lattice, we now calculate some invariants of the three-dimensional Ising model universality class.

3.2.1. Critical Exponents

According to the RG theory, critical exponents are determined entirely by a fixed point and do not depend on the lattice anisotropy. For a three-dimensional Ising model, the universality of critical exponents has been confirmed for $\Delta \in [0.2, 5]$ by the high-temperature series calculations [19].

At present, the most precise estimates of critical exponents are provided by the high-temperature expansions for ordinary models [20] and for models with

Table 2. Estimates of the thermal and magnetic critical exponents for different values of $\Delta = J'/J$. Calculations with a cluster pair (3, 4)

Δ	k_BT_c/J	y_t		y_h	
		Eq. (6)	Eq. (7)	Eq. (8)	Eq. (9)
1.0	4.51195062	1.5760695	1.7246286	2.5971647	2.5886128
0.5	2.92652189	1.5256373	1.6636718	2.5902006	2.5819462
0.1	1.34109829	1.4700811	1.5972576	2.5843305	2.5766511
0.05	1.03844933	1.4533899	1.5791439	2.5836720	2.5761101
0.01	0.65188600	1.4236178	1.5480583	2.5832982	2.5758028
0.005	0.55541801	1.4141719	1.5383503	2.5832029	2.5757888
0.001	0.40801006	1.3984754	1.5222765	2.5834573	2.5757953
		1.47{6}	1.60{7}	2.586{5}	2.579{5}

improved potentials characterized by suppressed leading scaling corrections [21]. For the three-dimensional (fully isotropic) Ising lattice, these methods yield $\nu = 0.63012(16)$ and $\gamma = 1.2373(2)$. Hence, $y_t = 1/\nu = 1.5870(4)$ and $y_h = (d + \gamma/\nu)/2 = 2.48180(18)$.

In Table 2, we report our estimates for the critical exponents y_t and y_h. It follows from those data that as the lattice anisotropy parameter Δ varies by three orders (from unity to 10^{-3}), the estimates of critical exponents are changed only by a few percent or less. In particular, calculations via Eqs. (6) and (7) with the cluster pair (3, 4) yield $y_t = 1.47\{6\}$ and $y_t = 1.60\{7\}$, respectively. (Here and below, the numbers in braces are dispersions of averages over the lattice anisotropy parameter Δ.) Their variations lie in the range 4–4.4%. Similar calculations of the magnetic critical exponent performed with Eqs. (8) and (9), also with the pair (3, 4), lead to $y_h = 2.586\{5\}$ and $y_h = 2.579\{5\}$, respectively. Relative dispersions of these estimates are about 0.2%.

Thus, our calculations confirm the universality of both critical exponents in an essentially wider range of Δ than in earlier investigations. Systematic errors of the estimates achieved arise due to small sizes L of the subsystems used.

3.2.2. Critical FSS Amplitude Ratios

Critical amplitudes are determined by scaling functions. As a result, their "universal ratios" like

$$\frac{A_{\kappa^{(4)}}}{A_{\chi^{(4)}}} = \frac{\mathcal{H}^{(0,4)}(0,0)}{\mathcal{F}^{(0,4)}(0,0)}$$

depend, generally speaking, on the lattice anisotropy because it can change the shape of subsystems. However, in the case of parallelepipeds $L^{d-1} \times \infty$ with unchanged (between themselves) transverse coupling constants, the shape of a sample (all its aspect ratios) is independent of the interaction in the longitudinal direction. Such a universality is studied here.

Table 3 contains results of our calculations for the critical FSS amplitude ratios $U = A_{\chi^{(4)}}A_\kappa/A_\chi^2$, $Y_1 = A_{\kappa''}/A_\chi$, and $Y_2 = A_{\kappa^{(4)}}/A_{\chi^{(4)}}$. Calculations have been performed for $\Delta \in [10^{-3}, 1]$ using a cyclic cluster $4 \times 4 \times \infty$.

In accordance with the data in Table 3, the average ratio $U = 4.900\{3\}$. Hence, as the anisotropy parameter Δ varies by three orders, this quantity changes only by 0.06%. With such accuracy, we may consider the given ratio a constant. In the case of a fully isotropic lattice, $A_\kappa = 1.26(5)$ and $A_{\chi^{(4)}}/A_\chi^2 = 3.9(2)$ [8], and therefore $A_{\chi^{(4)}}A_\kappa/A_\chi^2 = 4.9(5)$. Our values of U in Table 3 are in good agreement with this estimate.

It follows from Table 3 that $Y_1 = A_{\kappa''}/A_\chi = 1.759(2)$. Hence, the constancy of this universal amplitude ratio

Table 3. Estimates of the universal critical FSS amplitude ratios $U = A_{\chi(4)}A_\kappa/A_\chi^2$, $Y_1 = A_{\kappa(2)}/A_\chi$, and $Y_2 = A_{\kappa(4)}/A_{\chi(4)}$ for the Ising system with the cylindrical geometry $L \times L \times \infty$ and periodic boundary conditions. Data for $L = 4$

Δ	k_BT_c/J	U	Y_1	Y_2
1.0	4.51195062	4.8956599	1.7550004	2.0146443
0.5	2.92652189	4.8967625	1.7572512	2.0136519
0.1	1.34109829	4.9011909	1.7596003	2.0129829
0.05	1.03844933	4.9014406	1.7597697	2.0129285
0.01	0.65188600	4.9015375	1.7598563	2.0128977
0.005	0.55541801	4.9015529	1.7598646	2.0128953
0.001	0.40801006	4.9015782	1.7598732	2.0128938
		4.900{3}	1.759{2}	2.0133{6}

is estimated at least a few times 10^{-3}. Our average value for Y_1 agrees well with the estimate for the isotropic lattice, $A_{\kappa''}/A_\chi = 1.749(6)$ [8].

According to the data in Table 3, the amplitude ratio $Y_2 = A_{\kappa^{(4)}}/A_{\chi^{(4)}} = 2.0133\{6\}$. This quantity is therefore the most stable of all the invariants of the three-dimensional Ising universality class investigated in this paper. We note that we are not aware of any quantitative estimates for $A_{\kappa^{(4)}}/A_{\chi^{(4)}}$.

4. CONCLUSIONS

In this paper, large-scale transfer-matrix computations have been performed. Application of the extended phenomenological RG schemes has made it possible to find tight bounds on the critical temperature in the anisotropic simple-cubic Ising lattice and to improve the available estimates for it.

We calculated the thermal and magnetic critical exponents. Our results confirm the universality of y_t within 4–4.4% and of y_h within 0.2% over a remarkably wider range of Δ ($10^{-3} \le \Delta \le 1$) than in [19].

Finally, the results give clear evidence that the critical FSS amplitude ratios $U = A_{\chi^{(4)}}A_\kappa/A_\chi^2$, $Y_1 = A_{\kappa''}/A_\chi$, and $Y_2 = A_{\kappa^{(4)}}/A_{\chi^{(4)}}$ are independent of the lattice anisotropy parameter $\Delta = J'/J$ with accuracies of at least 0.1%. Probably for the first time in the literature, we give an estimate for the universal quantity Y_2.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project nos. 03-02-16909 and 04-03-32528).

REFERENCES

1. P. Nightingale, J. Appl. Phys. **53**, 7927 (1982).
2. M. N. Barber, in *Phase Transitions and Critical Phenomena*, Ed. by C. Domb and J. L. Lebowitz (Academic, New York, 1983), Vol. 8.
3. R. R. dos Santos and L. Sneddon, Phys. Rev. B **23**, 3541 (1981).
4. J. A. Plascak, W. Figueiredo, and B. C. S. Grandi, Braz. J. Phys. **29**, 579 (1999).
5. A. Pelissetto and E. Vicari, Phys. Rep. **368**, 549 (2002); cond-mat/0012164.
6. M. A. Yurishchev, Nucl. Phys. B (Proc. Suppl.) **83–84**, 727 (2000); Zh. Éksp. Teor. Fiz. **118**, 380 (2000) [JETP **91**, 332 (2000)].
7. M. A. Yurishchev, Phys. Rev. B **50**, 13533 (1994).
8. M. A. Yurishchev, Phys. Rev. E **55**, 3915 (1997).
9. M. A. Yurishchev, Zh. Éksp. Teor. Fiz. **125**, 1349 (2004) [JETP **98**, 1183 (2004)].
10. V. Privman and M. E. Fisher, Phys. Rev. B **30**, 322 (1984).
11. M. Hasenbusch, K. Pinn, and S. Vinti, Phys. Rev. B **59**, 11471 (1999).
12. H. W. J. Blöte, L. N. Shchur, and A. L. Talapov, Int. J. Mod. Phys. C **10**, 1137 (1999).
13. M. A. Novotny, Nucl. Phys. B (Proc. Suppl.) **20**, 122 (1991).
14. C. F. Baillie, R. Gupta, K. A. Hawick, and G. S. Pawley, Phys. Rev. B **45**, 10438 (1992).
15. C.-Y. Weng, R. B. Griffiths, and M. E. Fisher, Phys. Rev. **162**, 475 (1967); M. E. Fisher, Phys. Rev. **162**, 480 (1967).
16. T. Graim and D. P. Landau, Phys. Rev. B **24**, 5156 (1981).
17. M. A. Yurishchev and A. M. Sterlin, J. Phys.: Condens. Matter **3**, 2373 (1991).
18. R. Navarro and L. J. de Jongh, Physica B (Amsterdam) **94**, 67 (1978).
19. G. Paul and H. E. Stanley, Phys. Rev. B **5**, 2578 (1972).
20. P. Butera and M. Comi, Phys. Rev. B **62**, 14837 (2000); **65**, 144431 (2002).
21. M. Campostrini, A. Pelissetto, P. Rossi, and E. Vicari, Phys. Rev. E **60**, 3526 (1999); **65**, 066127 (2002).

Journal of Experimental and Theoretical Physics, Vol. 99, No. 3, 2004, pp. 544–551.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 3, 2004, pp. 625–633.

SOLIDS
Electronic Properties

Inverted Near-Surface Hysteresis Loops in Heterogeneous (Nanocrystalline/Amorphous) $Fe_{81}Nb_7B_{12}$ Alloys

E. E. Shalyguina[a,*], I. Škorvánek[b], P. Švec[c], V. A. Mel'nikov[a], and N. M. Abrosimova[a]

[a]*Moscow State University, Moscow, 119922 Russia*
[b]*Institute of Experimental Physics, Slovakian Academy of Sciences, SK-043 53 Košice, Slovakia*
[c]*Institute of Physics, Slovakian Academy of Sciences, SK-842 28 Bratislava, Slovakia*
**e-mail: shal@magn.ru*

Received January 22, 2004

Abstract—Inverted hysteresis loops were observed for the first time in the near-surface layers of heterogeneous (nanocrystalline/amorphous) $Fe_{81}Nb_7B_{12}$ alloys. In particular, a negative residual magnetization is observed when a positive magnetic field applied in the sample plane is decreased to zero. The inverted hysteresis is qualitatively explained within the framework of a two-phase model, according to which the heterogeneous alloys contain two dissimilar phases exhibiting uniaxial magnetic anisotropy and featuring antiferromagnetic exchange interaction. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In recent years, nanocrystalline materials of a new type were obtained by means of controlled crystallization of amorphous alloys [1, 2]. The new materials have attracted much interest because of their unique magnetic, mechanical, and kinetic properties. Iron-based Fe–M–B (M = Zr, Ta, Mo and Nb) alloys (NANOPERM) have proved to be most promising due to relatively simple composition and unique soft magnetic properties [2, 3], even in comparison with currently popular FeCuNbSiB alloys (FINEMET) [1]. The magnetic properties of Fe–M–B alloys (in particular, of the FeNbB system) have been studied by various experimental techniques (see, e.g., [2–10]). It was established that (i) FeNbB alloys annealed in the range from 200 to 800°C behave as materials containing two ferromagnetic phases (amorphous and nanocrystalline), (ii) the volume of the nanocrystalline phase (appearing upon annealing) depends on the method and temperature of treatment of the initial alloy, and (iii) the relative content of amorphous and nanocrystalline phases significantly influences the magnetic properties of the annealed material. For example, it was found [6] that the volume of a nanocrystalline phase in the $Fe_{80.5}Nb_7B_{12.5}$ annealed at T_{ann} = 490–550°C is small and the coercive force ($H_c \approx 2.6$ Oe) is almost two orders of magnitude higher as compared to that in the initial (unannealed) alloy; in the samples annealed at $T_{ann} \approx 600$°C, the volume of the nanocrystalline phase increases, and the H_c value decreases to 0.15 Oe. This change in the coercive force is caused by the appearance of a nanocrystalline phase in the annealed material.

Generally speaking, the excellent magnetic softness of nanocrystalline materials were explained by Herzer [11] within the framework of a model with random magnetic anisotropy. According to this model, nanocrystalline bcc Fe grains with a characteristic size smaller than the exchange length are strongly exchange-coupled via the amorphous matrix, which results in effective averaging of the magnetic anisotropy of individual nanocrystalline grains and in a significant decrease in magnetostriction which, in turn, leads to a considerable decrease in the magnetoelastic anisotropy and coercive force. Within the framework of this model, it is also possible to explain the above results. Indeed, in the FeNbB alloy annealed at ~490–550°C, the volume of a nanocrystalline phase is small and the nanocrystalline bcc Fe grains are separated by relatively thick layers of the amorphous magnetic matrix. As a result, the exchange interaction between nearest-neighbor bcc Fe grains is weak and the H_c is large. In the second case ($T_{ann} \approx 600$°C), the volume of the nanocrystalline phase is increased, the interaction between grains is enhanced and the coercive force is reduced.

An important role in the formation of magnetic properties of such materials is played by the surface. As is known, inhomogeneities in the microstructure and chemical composition of amorphous materials account for a significant (up to tenfold) increase in the surface coercive force H_c and the saturation field H_s as compared to the corresponding volume values [12, 13]. Moreover, these inhomogeneities significantly influence the course of crystallization in amorphous materials during their heat treatments used for the obtaining of

heterogeneous alloys containing two or more ferromagnetic phases.

The theory developed by Arrot [14] predicted the existence of inverted hysteresis loops in heterogeneous magnetic structures. A peculiarity of these loops consists in that a negative residual magnetization is observed a sample upon switching off the magnetic field applied to this sample in the positive direction. As the positive magnetic field decreases, the magnetization reversal in the sample in fact takes place under conditions of a positive coercive force H_c. It should be noted that the negative residual magnetization upon switching off the positive magnetic field is impossible (forbidden) in homogeneous magnetic materials, where the magnetization is the order parameter in description of their thermodynamic state. According to Arrot [14], this effect can take place only in heterogeneous magnetic systems.

Previously, inverted hysteresis loops have been experimentally observed in various inhomogeneous magnetic structures. In particular, this phenomenon takes place (i) in the Co–CoO system [15] and in cobalt films with a thickness of under 60 nm [16], where the inverted hysteresis loops were explained by the exchange anisotropy [17]; (ii) in multilayer Ag/Ni structures [18] and granular (Ni, Fe)–SiO_2 films [19], where the existence of such hysteresis loops was attributed to a magnetostatic interaction between magnetic layers and nanocrystalline grains [19, 20]; (iii) in epitaxial iron films, where the inverted hysteresis was related to the existence of a magnetization component perpendicular to the film plane [21]; (iv) in Co/Pt/Gd/Pt [22] and Fe/Au [23] superlattices, where the appearance of inverted hysteresis loops was explained in terms of the antiferromagnetic exchange interaction between Co and Cd (or Fe) layers through Pt (or Au) spacers; (v) and in CoNbZr and CoFeMoSiB films, YCo_2/YCo_2 bilayers, and FeNi/FeNi polycrystal samples [24, 25], where the inverted hysteresis loops were explained by simultaneous action of two noncollinear magnetic anisotropy fields on the average magnetization. These magnetic systems predominantly possess inhomogeneous layered structures with clear interfaces between layers. The interaction (magnetostatic [18–20] or exchange [15–17, 22, 23]) between layers in such systems accounts for the appearance of inverted hysteresis loops.

This paper reports the results of a comparative study of the surface and volume magnetic properties of heterogeneous $Fe_{81}Nb_7B_{12}$ alloys containing two (amorphous and nanocrystalline) magnetic phases. Taking into account that most significant difference in magnetic properties was observed for the samples of this FeNbB alloy annealed at 550 and 600°C, we restrict our consideration to the samples heat-treated at these temperatures. We will analyze how changes in the microstructure appearing upon heat treatment of the initial (unannealed) alloy affect the behavior of samples in the magnetic field.

2. SAMPLES AND EXPERIMENTAL METHODS

Amorphous $Fe_{81}Nb_7B_{12}$ ribbons with a width of 6 mm and a thickness of 30 μm were obtained by rapid solidification via melt spinning. The Ar-quenched ribbons were annealed in an argon atmosphere for 1 h at 550 or 600°C (referred to below as samples N1 and N2, respectively).

The crystal structure of samples was studied by X-ray diffraction. The measurements were performed using CuK_α radiation in the Bragg–Brentano geometry with a graphite monochromator in the reflected beam. The X-ray diffraction patterns were recorded at an angular step of $2\theta = 0.1°$; the parameters of a crystalline phase appearing upon annealing were determined using the results of scanning in the vicinity of the diffraction peaks at a step of 0.02°.

The near-surface magnetic properties of heterogeneous alloys were studied using a magnetooptical micromagnetometer described in detail elsewhere [26]. As is known, the magnetooptical Kerr effect is sensitive with respect to magnetization of the near-surface layer of a definite thickness corresponding to the light penetration depth t_{pen}. This layer thickness is determined by the formula $t_{pen} = \lambda/4\pi k$, where λ is the wavelength of the incident light and k is the absorption coefficient of the medium. According to the available experimental data [27], t_{pen} in ferromagnetic materials does not exceed 10–30 nm in the range of photon energies from 0.5 to 6 eV. In our case, the thickness of a probed near-surface layer was on the order of 20 nm.

The near-surface hysteresis loops were measured using equatorial (transverse) Kerr effect. The magnitude of this effect is determined as $\delta = (I - I_0)/I_0$, where I and I_0 are the intensities of light reflected from the magnetized and nonmagnetized sample, respectively. In fact, we studied the behavior of $\delta(H)/\delta_s \sim M(H)/M_s$ ratio (where δ_s is the value of the Kerr effect in the state of saturated magnetization and M_s is the saturation magnetization) in the course of cyclic variation of the magnetic field (from $+H$ to $-H$ and vice versa) applied parallel to the sample surface and perpendicular to the plane of light incidence. The anisotropy of magnetic properties was studied by rotating a sample around the normal to its surface. The angle between the direction along the ribbon length (the direction of spinning) and the orientation of magnetic field **H** is denoted by ϕ. The measurements were performed both on the free side of the ribbon and on the side contacting with the spinning wheel (commonly referred to as the wheel or contact side). The volume magnetic properties of samples were measured using a vibrating-sample magnetometer. All data presented below were obtained on round samples with a diameter of 6 mm cut from annealed ribbons,

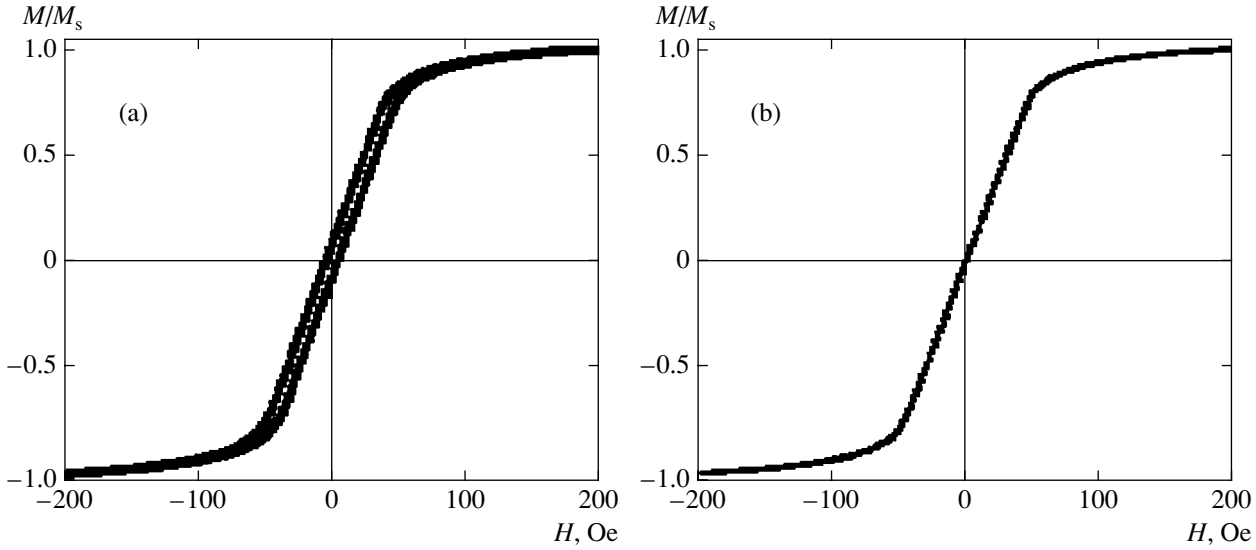

Fig. 1. The volume hysteresis loops observed in samples of (a) N1 (annealed at T_{ann} = 550°C) and (b) N2 (T_{ann} = 600°C) in the magnetic field applied in the sample plane and oriented at an angle $\phi = 0$ relative to the sample axis coinciding with the direction of melt spinning.

which eliminated the effect of shape anisotropy on the magnetic properties of samples.

3. RESULTS AND DISCUSSION

An analysis of the X-ray diffraction patterns showed that the initial (unannealed) $Fe_{81}Nb_7B_{12}$ alloy ribbon is amorphous, while the annealed samples of both types contain amorphous and nanocrystalline phases. The size of nanocrystalline bcc Fe grains formed in the samples annealed at T_{ann} = 550 and 600°C was 7–10 and 10–15 nm, respectively. It was found that the volumes of amorphous and nanocrystalline phases were 66.5 and 24.1% in the sample N1 and 53.9 and 38.6% in the sample N2, respectively. The volume of iron and niobium borides was 9.4% (type 1) and 7.5% (type 2). Thus, the volume of the nanocrystalline phase in the sample N2 was about one and half times that in type 1.

It was found that the aforementioned difference in the microstructure of samples annealed at the two temperatures was accompanied by significant differences in their magnetic behavior. Data obtained using the vibrating-sample magnetometer showed that the studied samples possess a slight in-plane magnetic anisotropy and magnetically soft properties. As is known [28], the magnetic behavior of amorphous alloys obtained by rapid solidification on a rotating wheel is determined by magnetoelastic anisotropy, which depends on the magnetostriction constant λ_s and the residual mechanical stresses σ induced along the ribbon in the course of melt spinning. The effective uniaxial magnetic anisotropy constant is defined by the relation $K \propto \lambda_s\sigma$, whereby the coercive force is $H_c \propto K/M_s$. Our samples were characterized by $\lambda_s \sim 10^{-6}$, which accounted for a weak magnetic anisotropy in the plane of the ribbon.

Figure 1 shows the volume hysteresis loops observed for annealed samples in the field **H** oriented at $\phi = 0°$. The loops exhibited the usual shape, and the values of the coercive force H_c^{vol} for the samples annealed at 550°C (type 1) and 600°C (type 2) were 4.3 and 0.1 Oe, respectively. According to the random magnetic anisotropy model generalized to the case of two-phase systems [13], an increase in the volume of bcc Fe grains in sample N2 leads to a more effective intergranular interaction, which accounts for a decrease in the coercive force.

It was established that the near-surface magnetic properties substantially differ from the volume characteristics. Figures 2 and 3 present the near-surface hysteresis loops measured in the magnetic fields oriented at $\phi = 0°$ and 90° for the samples N1 and N2, respectively. As can be seen, the shapes of the hysteresis loops observed for $\phi = 0°$ and 90° are significantly different for both samples, which is evidence of their strong surface magnetic anisotropy. No such anisotropy was observed in the as-cast sample (this case is illustrated in Fig. 4).

The surface coercive force H_c^{surf} is significantly higher than the volume value. For example, H_c^{surf} in sample N2 was almost two orders of magnitude higher than H_c^{vol}. This fact can be explained as follows. The results of examination of the cross sections of $Fe_{81}Nb_7B_{12}$ alloy ribbons in a transmission electron microscope using a technique described in [7] revealed a significant difference in microstructure of the volume

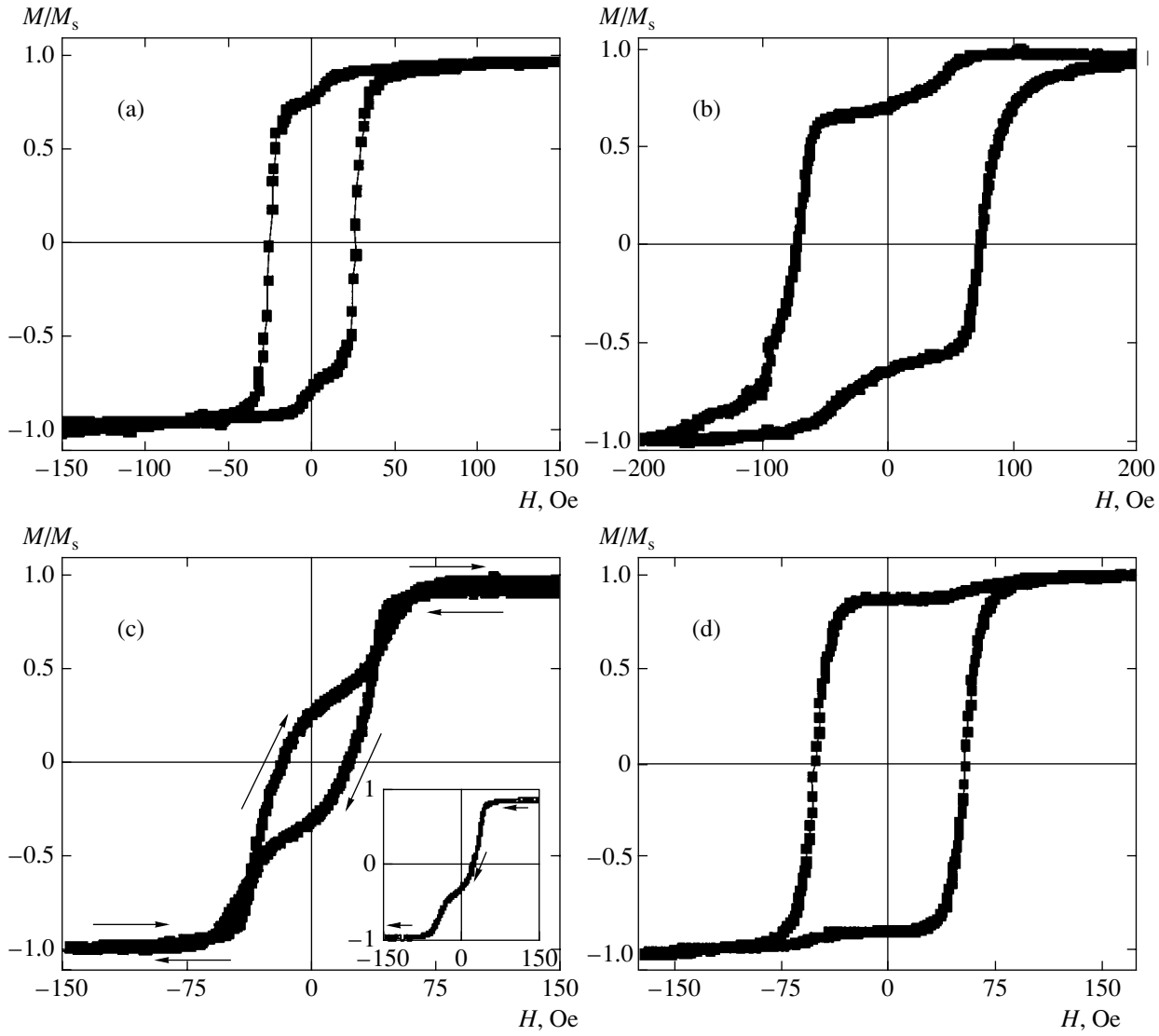

Fig. 2. The near-surface hysteresis loops observed for the (a, b) wheel and (c, d) free side of a sample N1 (T_{ann} = 550°C) in the magnetic field oriented at an angle of ϕ = 0° (a, c) and 90°(b, d). For the sake of clarity, the inset shows the forward branches of the loop.

and near-surface regions. In particular, it was found that the size of nanocrystalline bcc Fe grains varies from 50 to 120 and the concentration of nonmagnetic ions increases within an 0.4-μm-thick surface layer. These data are indicative of the presence of significant inhomogeneities in the microstructure and chemical composition in the surface layers of samples, which is just what accounts for a considerable increase in H_c^{surf} compared to H_c^{vol}. The large difference between H_c^{surf} and H_c^{vol} was observed virtually for all materials obtained by melt spinning [12, 13, 26].

It was also found that H_c^{surf} on the free side of ribbons is smaller than that on the wheel side. For example, on the free side of sample N2, H_c^{surf} = 12 and 46 Oe for ϕ = 0° and 90°, respectively, while the corresponding values on the wheel side are H_c^{surf} = 54 and 76 Oe. An important peculiarity of the method of manufacturing amorphous alloy ribbons by melt spinning consists in that the level of residual stresses developed on the wheel side of the ribbon is higher than on the free side; this difference is retained upon annealing ($\sigma_{wheel} > \sigma_{free}$. In addition, the wheel and free sides of amorphous alloy ribbons possess different surface morphologies. For example, according to atomic force microscopy data, the size of surface roughnesses on the free side does not exceed 2 nm, while that on the wheel surface reaches 10 nm. Thus, the differences in residual stresses developed on the wheel and free sides of the ribbon during manufacture, as well as in the surface morphology

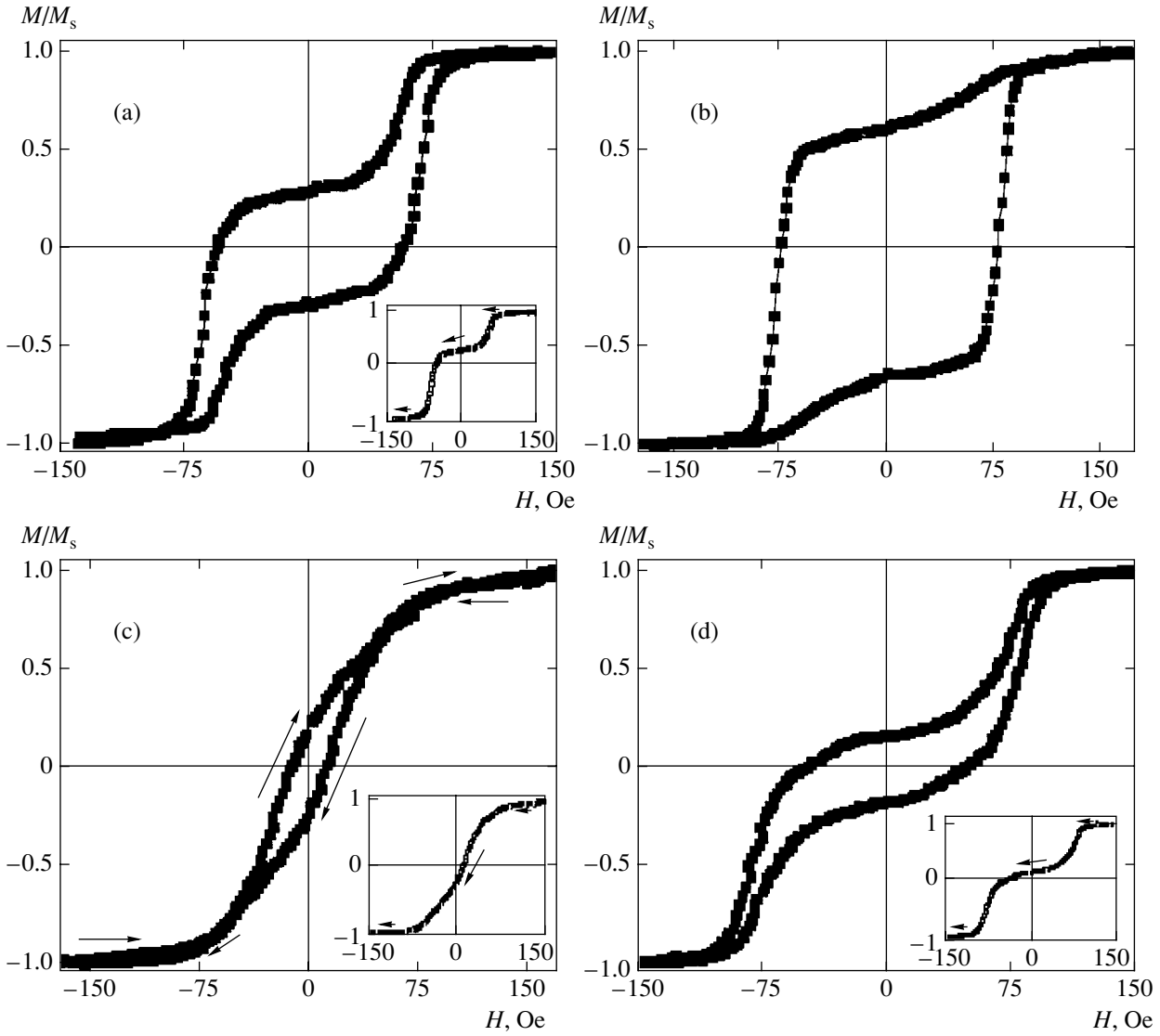

Fig. 3. The near-surface hysteresis loops observed for the (a, b) wheel and (c, d) free side of a sample N2 (T_{ann} = 600°C) in the magnetic field oriented at an angle of $\phi = 0°$ (a, c) and 90° (b, d). For the sake of clarity, the insets show the forward branches of the loop.

on the two sides, are factors accounting for the observed difference in the values of H_c^{surf}.

Of special importance is a rather complicated shape of the near-surface hysteresis loops observed in the annealed samples, which can be explained by their inhomogeneous magnetic structure, in particular, by the coexistence of amorphous and nanocrystalline phases characterized by different values of the magnetic anisotropy constants (K_1, K_2), saturation magnetizations (M_{s1}, M_{s2}), and coercive forces ($H_{c1,2} \propto K_{1,2}/M_{s1,2}$). As a result, the process of magnetization reversal in the two phases takes place in different magnetic fields, which accounts for the unusual behavior.

It is important to note that the two-phase character of annealed $Fe_{81}Nb_7B_{12}$ alloy samples is manifested only in the behavior of their near-surface layers featuring most significant changes in microstructure as a result of annealing. The volume magnetic characteristics of both samples also change in accordance with their microstructure, but the shape of the hysteresis loop exhibits no significant variations.

A more detailed analysis of the experimental data presented in Figs. 2c and 3c showed that the forward and reverse branches of the near-surface hysteresis loops measured for $\phi = 0$ exhibit unusual behavior in response to the field variation. In particular, a negative residual magnetization is observed when the positive magnetic field decreases to zero and the magnetization reversal takes place at a positive value of the coercive force (that is, the hysteresis loop are inverted). This fact is clearly illustrated in the insets to Figs. 2c and 3c

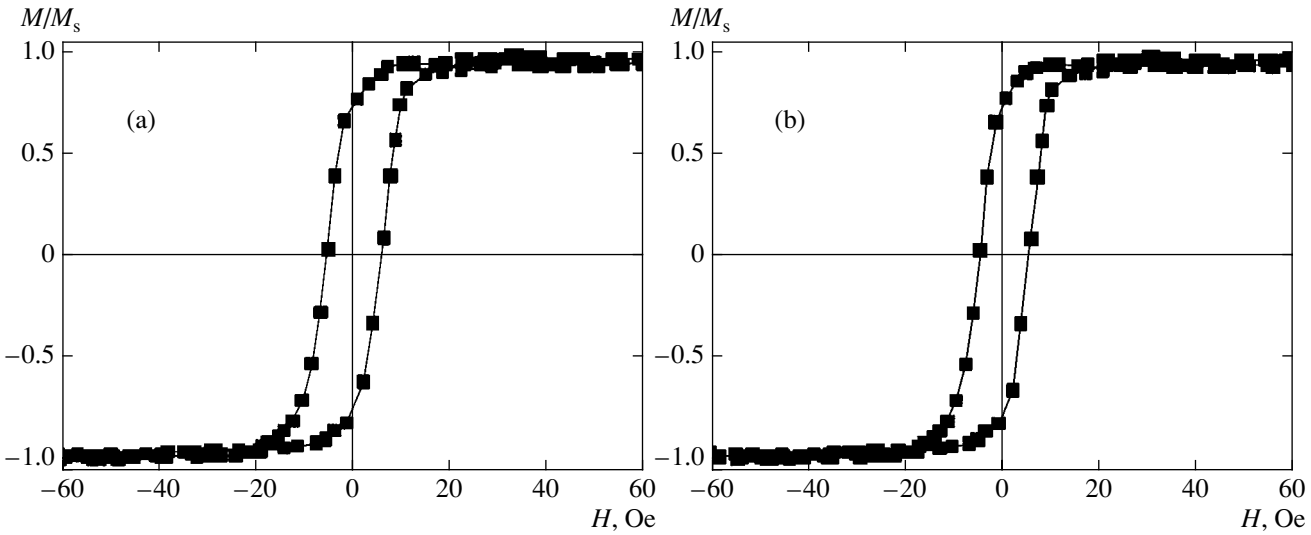

Fig. 4. The near-surface hysteresis loops observed for the wheel side of an as-cast sample in the magnetic field oriented at an angle $\phi = 0°$ (a) and 90° (b). The surface coercive force is about 26 times the volume value.

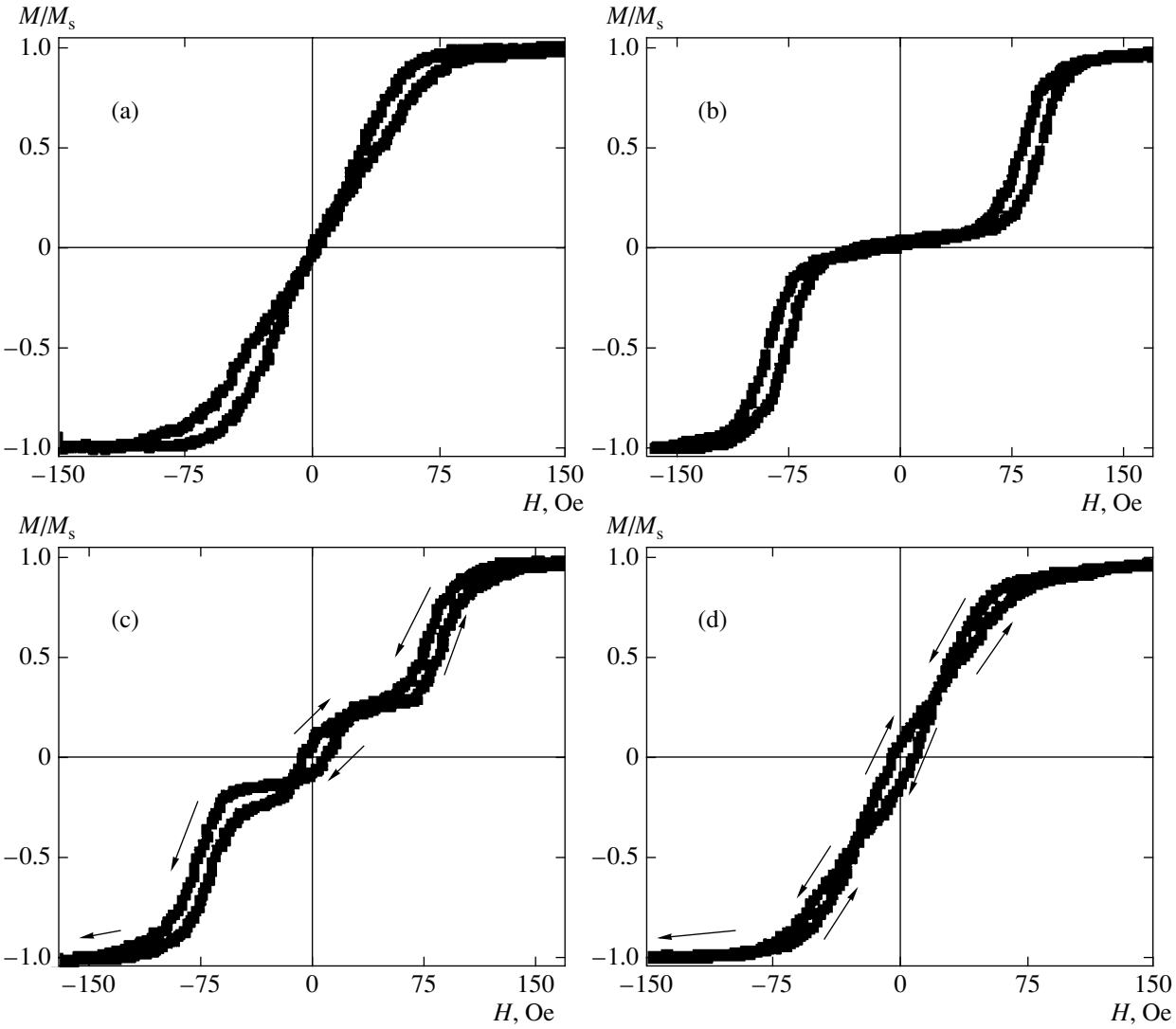

Fig. 5. The near-surface hysteresis loops observed for the free side of a sample N2 ($T_{ann} = 600°C$) in the magnetic field oriented at an angle $\phi = 2°$ (a), 105° (b), 135° (c), and 150° (d).

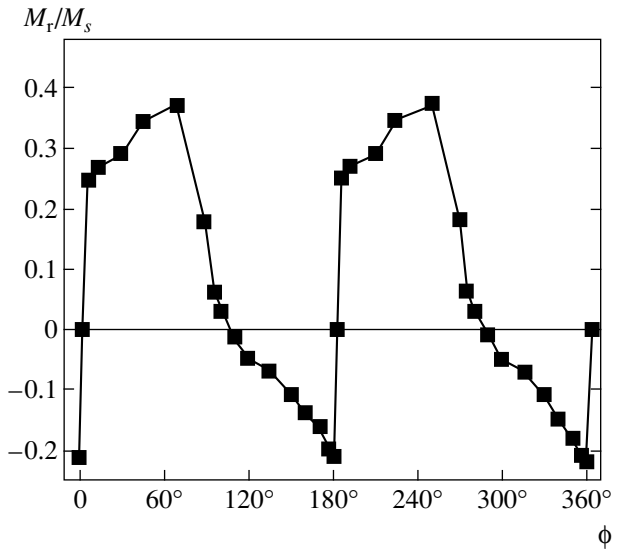

Fig. 6. A plot of the reduced residual magnetization versus angle ϕ for the free side of a sample N2 (T_{ann} = 600°C).

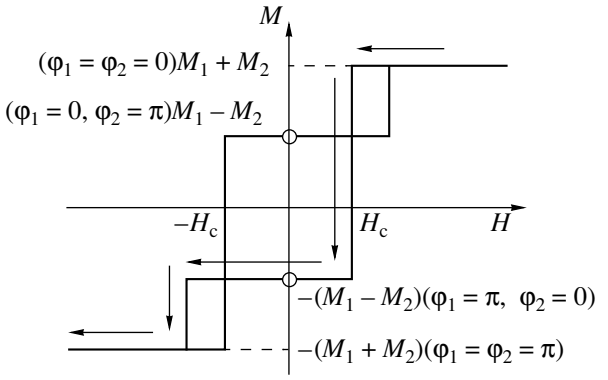

Fig. 7. A hysteresis loop calculated using a model of two nonidentical phases with uniaxial magnetic anisotropy and antiferromagnetic exchange.

showing only one branch of the hysteresis loop (observed for the field varied from $+H$ to $-H$).

In order to elucidate this phenomenon, we measured the near-surface hysteresis loops for various direction of the external magnetic field, whereby the angle ϕ was varied from 0° to 360° at a 5° step. The results of these measurements showed that the shape of the near-surface hysteresis loops strongly depends on the orientation of magnetic field in the sample plane. This is illustrated in Fig. 5, which shows the most significantly different shapes of the hysteresis loop observed on the free side of the sample N2. An analysis of these data revealed unusual behavior of the residual magnetization as a function of the field orientation angle ϕ. Figure 6 shows the dependence of the reduced residual magnetization M_r/M_s on the ϕ value for the free side of sample N2. As can be seen, there is an interval of angles where M_r/M_s is negative. It was established that the near-surface hysteresis loops measured for such ϕ values are completely (see Figs. 2c, 3c) or partly (see Fig. 5) inverted. Analogous dependences of M_r/M_s on the angle ϕ were observed for sample N2. It should be noted that the initial amorphous $Fe_{81}Nb_7B_{12}$ alloy also exhibited a certain difference between the volume and near-surface hysteresis loops (e.g., $H_c^{surf} \approx 5$ Oe against $H_c^{vol} \approx 0.2$ Oe on the wheel side at ϕ = 0), but the shapes of the bulk and near-surface hysteresis loops were virtually identical and the inverted loops were not observed.

The above experimental data can be qualitatively explained within the framework of a two-phase model according to which the heterogeneous alloys under consideration contain two dissimilar phases exhibiting uniaxial magnetic anisotropy and featuring antiferromagnetic exchange interaction. The total free energy of this system can be expressed as

$$E = -M_1 H\cos\varphi_1 - M_2 H\cos\varphi_2 + K_1\sin^2(\varphi_1 - \beta_1) + K_2\sin^2(\varphi_2 - \beta_2) + 2J_{12}M_1M_2\cos(\varphi_1 - \varphi_2), \quad (1)$$

where the first two terms represent the Zeeman magnetic energies of two phases with magnetizations M_1 and M_2; the third, fourth, and fifth terms describe the anisotropic energies of these phases and the exchange interaction between them (K_1 and K_2 are the constants of uniaxial magnetic anisotropy and J_{12} is the constant of antiferromagnetic exchange interaction); φ_1 and φ_2 are the angles between magnetization vectors M_1 and M_2 are the field direction; and β_1 and β_2 are the angles between the magnetic anisotropy axes of the corresponding phases and the field direction (the magnetic field **H** is oriented in the sample plane). The magnetization of this system is given by the formula

$$M = M_1\cos\varphi_1 + M_2\cos\varphi_2. \quad (2)$$

The values of angles φ_1 and φ_2 as functions of the magnetic field H can be determined from the system of equations

$$\frac{\partial E}{\partial\varphi_1} = \frac{\partial E}{\partial\varphi_2} = 0. \quad (3)$$

Assuming that the anisotropic energies are much greater than the exchange interaction energy, $\beta_1 = \beta_2 = 0$, and $M_1 > M_2$ and taking into account the condition $\partial^2E/\partial\varphi_1^2 = \partial^2E/\partial\varphi_2^2 > 0$ determining stable solutions of Eqs. (3), we calculated the hysteresis loop (Fig. 7). As can be seen, the simplified two-phase model explains only the partly inverted hysteresis loops. In order to obtain the completely inverted hysteresis, it is neces-

sary to perform the calculations without any simplifying assumptions, which is a rather difficult task for the system under consideration.

4. CONCLUSIONS

We have studied the surface and volume magnetic properties of heterogeneous (nanocrystalline/amorphous) $Fe_{81}Nb_7B_{12}$ alloys and revealed a strong influence of structural changes in annealed samples on their surface magnetic characteristics. Completely and partly inverted hysteresis loops in these materials were observed for the first time. A positive coercive force and a negative residual magnetization were observed when the positive magnetic field was decreased to zero. The experimental results were qualitatively explained within the framework of a two-phase model, according to which the heterogeneous alloys contain two dissimilar phases characterized by uniaxial magnetic anisotropy and antiferromagnetic exchange interaction.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research, project no. 02-02-16627.

REFERENCES

1. Y. Yoshizawa, S. Oguma, and K. Yamauchi, J. Appl. Phys. **64**, 6044 (1988).
2. K. Suzuki, A. Makino, A. Inoue, and T. Masumoto, J. Appl. Phys. **74**, 3316 (1993).
3. A. Makino, T. Hatanai, A. Inoue, and T. Masumoto, Mater. Sci. Eng. A **226–229**, 594 (1997).
4. K. Suzuki, A. Makino, A. Inoue, and T. Masumoto, J. Appl. Phys. **70**, 6232 (1991).
5. K. Suzuki, N. Kataoka, A. Inoue, *et al.*, Mater. Trans., JIM **31**, 743 (1990).
6. M. Kopcewicz, A. Grabias, I. Škorvánek, *et al.*, J. Appl. Phys. **85**, 4427 (1999).
7. M. Miglierini, M. Kopcewicz, B. Idzikowski, *et al.*, J. Appl. Phys. **85**, 1014 (1999).
8. I. Škorvánek, C. G. Kim, J. Kováč, *et al.*, J. Magn. Magn. Mater. **215–216**, 440 (2000).
9. I. Škorvánek, S. Skwirblies, and J. Kötzler, Phys. Rev. B **64**, 184437 (2001).
10. I. Škorvánek, P. Švec, J. M. Grenèche, *et al.*, J. Phys.: Condens. Matter **14**, 4717 (2002).
11. G. Herzer, Phys. Scr. **49**, 307 (1993).
12. E. E. Shalyguina, L. M. Bekoeva, and N. I. Tsidaeva, Sens. Actuators A **81**, 216 (2000).
13. A. Hernando, M. Vázquez, T. Kulik, and C. Prados, Phys. Rev. B **51**, 3581 (1995).
14. A. S. Arrot, in *Nanomagnetism*, Ed. by A. Hernando (Kluwer, Dordrecht, 1993), p. 73.
15. M. J. O'Shea and A. L. Al-Sharif, J. Appl. Phys. **75**, 6673 (1994).
16. P. W. Haycock, M. F. Chioncel, and J. Shah, J. Magn. Magn. Mater. **242–245**, 1057 (2002).
17. J. Nogués and I. K. Schuller, J. Magn. Magn. Mater. **192**, 203 (1999).
18. C. A. dos Santos and B. Rodmacq, J. Magn. Magn. Mater. **147**, L250 (1995).
19. X. Yan and Y. Xu, J. Appl. Phys. **79**, 6013 (1996).
20. A. Aharoni, J. Appl. Phys. **76**, 6977 (1994).
21. M. Cougo dos Santos, J. Geshev, J. E. Schmidt, *et al.*, Phys. Rev. B **61**, 1311 (2000).
22. K. Takanashi, H. Kurokawa, and H. Fujimori, Appl. Phys. Lett. **63**, 1585 (1993).
23. T. Šlezak, W. Karaś, K. Krop, *et al.*, J. Magn. Magn. Mater. **240**, 362 (2002).
24. S. M. Valvidares, L. M. Álvarez-Prado, J. I. Martin, and J. M. Almeda, Phys. Rev. B **64**, 134423 (2001).
25. S. M. Valvidares, J. I. Martin, L. M. Álvarez-Prado, *et al.*, J. Magn. Magn. Mater. **242–245**, 169 (2002).
26. E. E. Shalyguina, V. V. Molokanov, and M. A. Komarova, Zh. Éksp. Teor. Fiz. **122**, 593 (2002) [JETP **95**, 511 (2002)].
27. E. E. Shalyguina and K. H. Shin, J. Magn. Magn. Mater. **220**, 167 (2000).
28. K. Suzuki, H. Fujimori, and K. Hashimoto, *Amorphous Metals*, Ed. by T. Masumoto (Tokyo, 1982).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 3, 2004, pp. 552–558.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 3, 2004, pp. 634–641.

SOLIDS
Electronic Properties

Excitation of Bloch Oscillations in a Lateral Semiconductor Superlattice under the Influence of Electromagnetic Pulses

E. P. Dodin[a], A. A. Zharov[a], and A. M. Malkin[b]

[a]*Institute for Physics of Microstructures, Russian Academy of Sciences, Nizhni Novgorod, 603950 Russia*
[b]*Institute of Applied Physics, Russian Academy of Sciences, Nizhni Novgorod, 603950 Russia*
e-mail: dodin@ipm.sci-nnov.ru
Received November 5, 2003

Abstract—The conversion of the carrier frequency of electromagnetic pulses in lateral semiconductor superlattices, associated with the excitation of Bloch oscillations in the superlattice, is studied theoretically. Conditions are found that are necessary for the observation of the radiation of a Bloch oscillator. The energy characteristics of the efficiency of frequency multiplication and the spectral distribution of the radiation transmitted through the superlattice are calculated. It is shown that low-frequency collisions of electrons do not suppress the excitation of Bloch oscillations, which can be observed under the interaction of the superlattice not only with a pulsed, but also with a continuous-wave signal. © *2004 MAIK "Nauka/Interperiodica".*

1. One of the main factors that hampers the practical implementation of terahertz sources of electromagnetic radiation that are based on the resonance properties of a Bloch oscillator is the low-frequency instability of space-charge waves. This instability stems from the fact that the real part of the small-signal differential conductivity of a semiconductor superlattice is negative even at zero frequency [1]. As a result, the total electric field in the superlattice is decomposed into small-scale drifting domains, giving rise to Gunn-type oscillations of relatively low frequency. This leads to the suppression of the total negative conductivity of the superlattice placed in a dc electric field, near the Bloch frequency $\omega_B = eEd/\hbar$, where E is the intensity of the applied electric field, e is the electron charge, and d is the period of the superlattice potential, due to the bunching of electrons in the momentum space [2]. There are a number of theoretical studies in which the authors discuss various methods for suppressing the domain instability. Among these methods are those that are associated, in particular, with the attainment of a strongly nonlinear stage of high-frequency instability near the Bloch resonance [3, 4], where the dc differential conductivity of the superlattice becomes positive, or with the injection of hot electrons into the upper part of the miniband [5], which shifts the excitation band to frequencies slightly greater than the Bloch frequency, where the domain instability is suppressed. However, the implementation of these ideas is highly complicated by a number of factors both of technical and fundamental natures.

It was established about ten years ago [6, 7] that semiconductor superlattices are characterized by a superfast response to external electric fields. The character of this response depends on the intensity of the electric field applied to a superlattice and represents either an aperiodic relaxation to the ascending (stable) branch of the Esaki–Tsu current–voltage characteristic [8] for an electric-field intensity less than a certain critical value corresponding to the Bloch frequency, which is equal to the scattering frequency of electrons, or a quasi-periodic process with a frequency close to the frequency of Bloch oscillations of electrons in the case of supercritical fields. The relaxation time of the response is largely determined by the inverse of the collision frequency of electrons and, depending on the regularity of the structure and the operating temperature, amounts to several tens of picoseconds in real superlattices. Thus, one can expect that the irradiation of a superlattice by short electromagnetic pulses of length comparable with the inverse collision frequency of electrons, which guarantees a coherent motion of charge carriers during the whole period of interaction, leads to the excitation of Bloch oscillations. It should be emphasized that domains are not generated in the absence of a dc bias field because we deal with frequencies of terahertz and subterahertz bands, which, as a rule, are greater than the increment of the domain instability. The frequency of Bloch oscillations depends on the intensity of the instantaneous self-consistent electric field in the superlattice and may be much greater than the frequency of the incident wave. This fact opens up possibilities not only for the direct observation of Bloch oscillations but also for the multiplication of the carrier frequency of the incident pulse. It is the determination of conditions under which Bloch oscillations arise in the superlattice under the irradiation by an electromagnetic pulse and of the spectral distribution of the

radiation transmitted through the superlattice that is the subject of the present study. In recent works [9–11], it was shown that lateral semiconductor superlattices in which a two-dimensional electron gas is localized near the surface of a substrate and the miniband energy spectrum is formed under the motion of electrons along this surface represent the most suitable objects that have a highly regular structure [12, 13] and allow the observation of Bloch oscillations and frequency multiplication.

2. Consider a lateral superlattice of thickness h placed on a dielectric substrate of permittivity ε_s (Fig. 1). The interaction between the superlattice and an electromagnetic wave that is normally incident to the superlattice and is polarized so that the electric field is perpendicular to the layers that form the superlattice is described by the following system of equations, which was first derived in [9] (see also [10, 11]):

$$\frac{dV}{dt} = \frac{e}{m(W)}E - \nu_V V, \tag{1a}$$

$$\frac{dW}{dt} = eEV - \nu_W(W - W_T), \tag{1b}$$

$$E_{inc}(t) = E(t) + \frac{2\omega_p^2\hbar^2\mu_0 h}{e\Delta d^2 c(1+\sqrt{\varepsilon_s})}V(t). \tag{1c}$$

Here, V and W are the average (hydrodynamic) values of the velocity and energy of electrons, ν_W is the frequency of inelastic collisions, $\nu_V = \nu_W + \nu_{el}$ is the relaxation frequency of velocity (ν_{el} is the frequency of elastic collisions), $W_T = (\Delta/2)(1 - \mu_0)$ is the mean thermal energy of electrons in the absence of electric field, $\mu_0 = I_1(\Delta/2kT)/I_0(\Delta/2kT)$, T is temperature, k is the Boltzmann constant, $I_{0,1}(x)$ are the modified Bessel functions, Δ is the miniband width, $m(W) = (2\hbar^2/\Delta d^2)(1 - 2W/\Delta)^{-1}$ is the electron effective mass, d is the period of the lateral superlattice, E is the intensity of the electric field in the superlattice, $E_{inc}(t)$ is the intensity of the electric field incident on the surface of the superlattice, $\omega_p^2 = 4\pi e^2 n_e/m(W = 0)$ is the plasma frequency of electrons at the bottom of the miniband, and n_e is the electron concentration. The constitutive equations (1a) and (1b) describe electron transport in the lateral superlattice in a one-miniband quasi-classical approximation [6, 14]. Equation (1c) determines a relation between the intensity of the self-consistent electric field in the superlattice and the intensity of the field in the incident electromagnetic wave (a detailed derivation is given in [9, 11]) for a thin (compared with the wavelength) superlattice (simple calculations show that this condition is certainly fulfilled in real structures in the terahertz band). In this case, the electric field is uniform along the thickness of the superlattice. This allows one to replace the volume currents through the superlattice

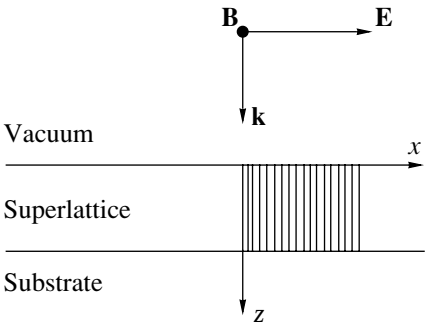

Fig. 1. Geometry of the problem.

by an equivalent surface current that determines the discontinuity of the tangential component of the magnetic field. One should also note that, within the approximations used, the intensity of the electric field in the lateral superlattice coincides with the intensity of the electric field in the wave transmitted through the superlattice.

Under the conditions of quasi-ballistic transport of electrons in a lateral superlattice, which may occur either in the absence of collisions for $\nu_W = \nu_V = 0$ or, as was pointed out above, under high-frequency pumping by sufficiently short electromagnetic pulses of about $\tau \sim 1/\nu_{V,W}$, the system of equations (1a), (1b) has the following integral of motion:

$$\left(\frac{2\hbar}{\mu_0 d\Delta}\right)^2 V^2(t) + \left[1 - \frac{2[W(t) - W_T]}{\mu_0\Delta}\right]^2 = \text{const}.$$

Assuming that $V(t_0) = 0$ and $W(t_0) = W_T$ at a certain moment t_0 (and thus choosing a phase trajectory corresponding to zero initial conditions for the velocity of electrons and the deviation of their energy from the equilibrium value), we find that const = 1. As a result, the expressions for the velocity and energy of electrons can be rewritten as

$$V(t) = \frac{\mu_0 d\Delta}{2\hbar}\sin A(t),$$
$$W(t) = W_T + \frac{\mu_0\Delta}{2}[1 - \cos A(t)]. \tag{2}$$

Substitution of (2) into (1a) yields

$$E(t) = \frac{\hbar}{ed}\frac{dA(t)}{dt}, \tag{3}$$

where $A(t)$ has the meaning of the dimensionless vector potential. Combining Eqs. (1c), (2), and (3), we can reduce the system of equations (1) with the initial conditions given above to a single equation

$$\frac{dA}{d\tau} + \theta\sin A = u_i(\tau) \tag{4}$$

with the initial condition $A(t = t_0) = 0$. Here,

$$\theta = \frac{\omega_p^2 h\mu_0}{\omega c(1+\sqrt{\varepsilon_s})}, \quad \tau = \omega t, \quad u_i(\tau) = \frac{eE_{inc}(\tau)d}{\hbar\omega},$$

and ω is the circular frequency of the wave incident to the lateral superlattice. Equation (4) coincides with the equation of motion of an overdamped pendulum (in this case, the role of the deviation angle is played by the vector potential A) under the action of an alternating moment of forces.[1] Thus, one may expect that the character of interaction between the electromagnetic wave and the lateral superlattice described by Eq. (4) essentially depends on the peak intensity of the incident pulse, which corresponds (by a mechanical analogy) to two types of dynamic behavior of the pendulum, without and with rotation. In this case, the rotation of the pendulum is associated with the excitation of Bloch oscillations in the superlattice. Indeed, consider a solution to Eq. (4) with the time-independent right-hand side $u_i(\tau) = u_0 = \text{const} > \theta$, which can be expressed as

$$A(\tau) = 2\arctan\left\{\frac{\theta}{u_0} + \sqrt{1-\frac{\theta^2}{u_0^2}}\tan\left(\frac{u_0}{2}\sqrt{1-\frac{\theta^2}{u_0^2}}(\tau-\tau_1)\right)\right\}, \quad (5)$$

$$u(\tau) = \frac{dA}{d\tau} = \frac{u_0^2-\theta^2}{u_0+\theta\cos(\sqrt{u_0^2-\theta^2}(\tau-\tau_2))}, \quad (6)$$

where

$$\tau_1 = \omega t_0 + \frac{2}{\sqrt{u_0^2-\theta^2}}\arctan\frac{\theta}{\sqrt{u_0^2-\theta^2}},$$

$$\tau_2 = \tau_1 + \frac{1}{\sqrt{u_0^2-\theta^2}}\arccos\sqrt{1-\frac{\theta^2}{u_0^2}},$$

and $u(\tau)$ is a dimensionless electric field in the superlattice. Solution (5), (6) describes anharmonic periodic oscillations of the electric field in the lateral superlattice[2] at the fundamental frequency $\Omega = \sqrt{u_0^2-\theta^2}$, which coincides with the eigenfrequency of a radiatively damped Bloch oscillator. In contrast to the collision mechanism of attenuation, which leads to the phase mixing of electrons that oscillate at the Bloch frequency and, as a consequence, to the attenuation of the macroscopic oscillatory response, the radiation mechanism of dissipation does not break the phase coherence of electron oscillations and does not lead to the attenuation of Bloch oscillations in the superlattice in a dc electric field. At the same time, any energy-dissipation mechanism leads to the shift of the oscillator eigenfrequency, which is determined in this case by the flux of electromagnetic energy, emitted into the surrounding medium, which is proportional to the surface concentration of electrons, i.e., to the parameter θ.

3. It is fairly obvious that the generation of high-frequency oscillations also occurs in the case of an ac incident field. The interaction between short pulses and a lateral superlattice, which leads to the excitation of Bloch oscillations, can be qualitatively represented as follows. As long as the instantaneous intensity of the electric field in the incident wave is less than the critical value, $u_i(\tau) < u_c = \theta$, the magnitude of the self-consistent field in the superlattice is small (the superlattice is in the shielding state), and the generation of harmonics is insignificant. At the moments when $u_i(\tau) > u_c$, the superlattice becomes transparent and Bloch oscillations of current frequency $\Omega = \sqrt{u_i^2(\tau)-\theta^2}$ are generated due to the increase in the intensity of the self-consistent electric field in the superlattice. Since the frequency of Bloch oscillations is determined by the amplitude of the incident field, it may be substantially greater than the frequency of the pumping wave. Thus, one may expect that the spectrum of the wave transmitted through the lateral superlattice, which arises due to multiplication of the incident-wave frequency as a result of excitation of Bloch oscillations, should become noticeably richer.

This interaction scenario is illustrated in Fig. 2, which presents the results of the numerical integration of Eq. (4) for an incident pulse with a Gaussian envelope,

$$u_i(\tau) = u_0(\tau)\sin\tau, \quad u_0(\tau) = u_m\exp\{-(\tau-T_0)^2/T_0^2\}, \quad (7)$$

where T_0 is the characteristic length of the pulse. Notice the slightly "above-threshold" regime $u_m - \theta \ll \theta$ (Fig. 2d) under which, according to calculations, a situation can be realized when only one half-period of Bloch oscillations is generated during one half-period of the incident field. This regime can be used for generating ultrashort electromagnetic pulses of length $\tau_p \ll 1$. Figure 3 shows the numerically calculated characteristic spectra of the radiation transmitted through the superlattice under conditions when the peak values of the intensity of the incident pulse are greater than the critical value. One can see a group of higher order harmonics in the spectra that correspond to the excitation of Bloch oscillations in the superlattice. At the same time, the higher order harmonics in appropriate spectra provide evidence for the multiplication of the radiation frequency.

4. In some special cases, the energy efficiency of the frequency conversion can be evaluated analytically. Consider incident pulses with a rectangular envelope function: $u_i(\tau) = u_m\sin\tau$ for $0 < \tau < T_0$ and $u_i(\tau) = 0$ oth-

[1] It should be noted that Eq. (4) also describes the dynamics of the phase difference of the order parameter in a Josephson junction with small capacitance [15].

[2] When $u_0 < \theta$, there are no periodic solutions.

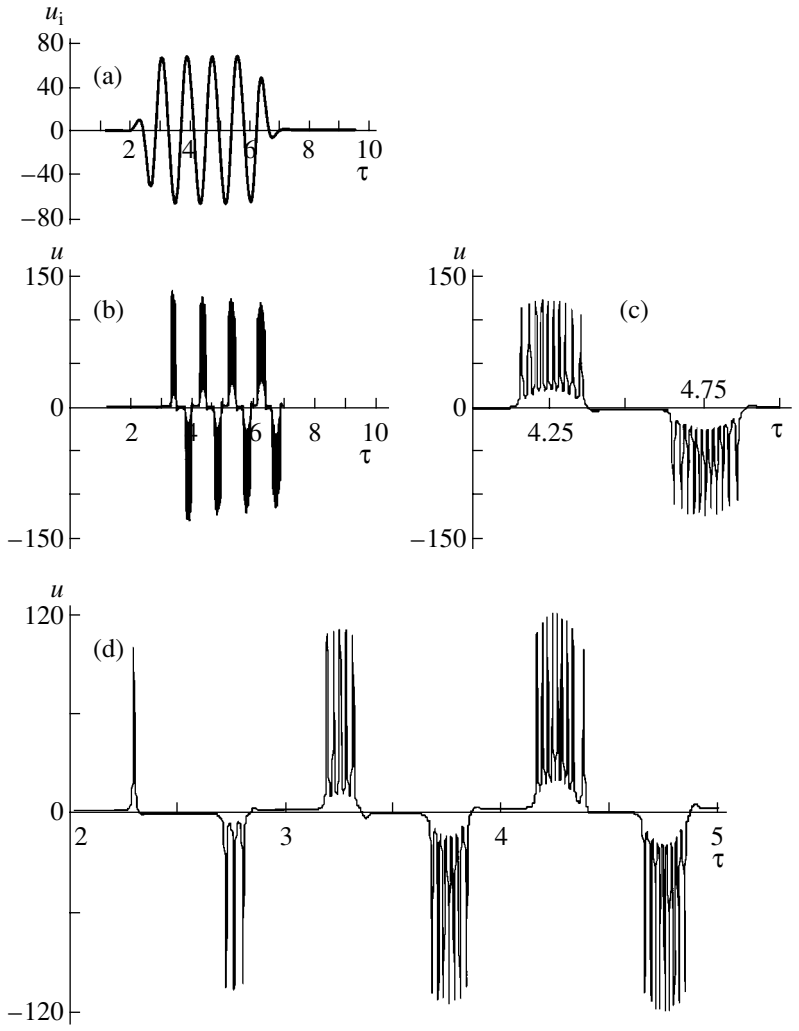

Fig. 2. (a) Oscillogram of the electric-field intensity in a pulse with a Gaussian envelope incident to the lateral superlattice. (b) Oscillogram of the electric field intensity in the pulse transmitted through the superlattice. (c) The fine structure of one period of the field transmitted through the superlattice. (d) Oscillogram of the field transmitted through the superlattice: an example of the situation when only one half-period of Bloch oscillations is generated within one half-period of the incident field.

erwise. As a rough estimate, assume that, when $|u_i(\tau)| < \theta$, the lateral superlattice completely shields the radiation (the electric-field intensity in the superlattice is equal to zero), whereas, when the inverse inequality holds, the solution can be represented as (6), which actually corresponds to a model in which the incident field (above a certain threshold) is constant and equal to its amplitude value u_m.

Calculating the amplitude of the electric field at the fundamental frequency by the Fourier transform technique, we obtain

$$u_\Omega = -2\theta\frac{\sqrt{M^2-1}}{M+\sqrt{M^2-1}}, \qquad (8)$$

where $M = u_m/\theta$. Next, calculating the ratio of the energy irradiated from the lateral superlattice at frequency Ω during a half period of the incident field (per unit surface of the superlattice) to the amount of energy incoming to the superlattice by the incident electromagnetic wave, we arrive at the following expression for the energy efficiency of frequency conversion in the lateral superlattice:

$$K = \frac{8}{\pi}\frac{M^2-1}{M^2}\frac{1}{(M+\sqrt{M^2-1})}\arccos\frac{1}{M}. \qquad (9)$$

Function (9) is plotted in Fig. 4. One can see that the efficiency of frequency multiplication attains its maximum of $K = 0.175$ at $M \approx 1.5$. In this case, the characteristic frequency of Bloch oscillations is $\Omega \approx 1.25\theta$. Thus, for sufficiently large θ, not only the frequency multiplication ratio but also the efficiency of frequency conversion may reach large values.

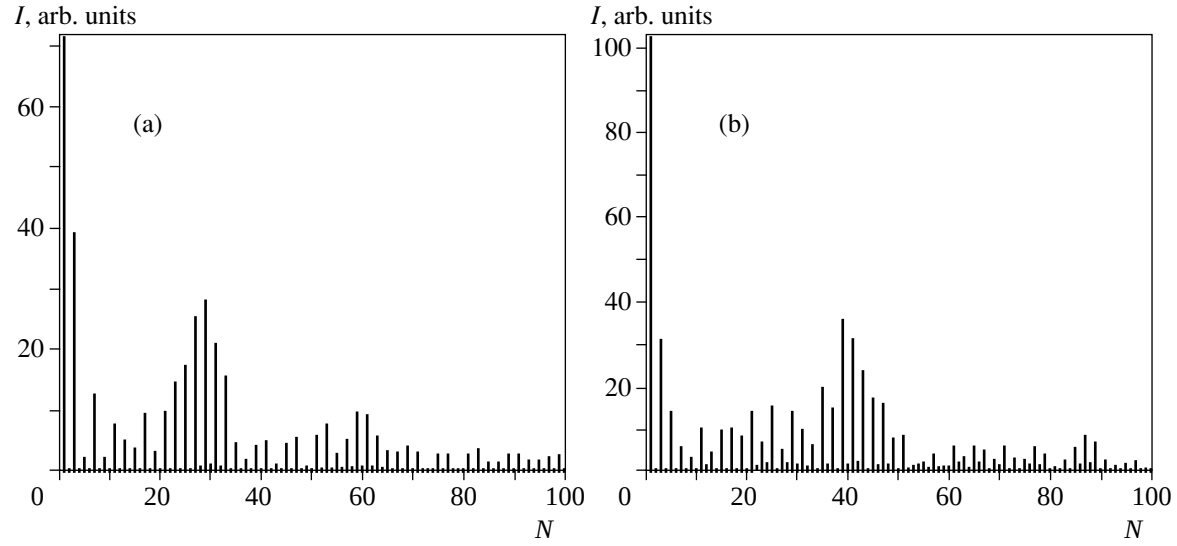

Fig. 3. Spectra of the radiation transmitted through the lateral superlattice; (a) $\theta = 30$ and $u_m = 44$ and (b) $\theta = 30$ and $u_m = 54$.

5. The effect of low-frequency collisions of electrons can be evaluated by perturbation theory. To this end, we write the system of equations (1) in dimensionless variables as follows:

$$\frac{dw}{d\tau} = u(1-\zeta) - \nu_1 w, \tag{10a}$$

$$\frac{d\zeta}{d\tau} = wu - \nu_2 \zeta, \tag{10b}$$

$$u_i(\tau) = u + \theta w, \tag{10c}$$

where

$$w = \frac{2\hbar V}{d\mu_0 \Delta}, \quad \zeta = \frac{2(W - W_T)\mu_0}{\Delta}, \quad \nu_{1,2} = \frac{\nu_{V,W}}{\omega}.$$

The other dimensionless parameters entering (10) have been introduced above. Without detriment to the physical content, we neglect elastic collisions, so that $\nu_1 = \nu_2 = \nu$. We also assume that the collision frequency is much less than the carrier frequency of the electromagnetic pulse incident to the superlattice ($\nu \ll 1$). Before applying perturbation theory, it is convenient to rewrite system of equations (10) in the complex variable $Z = \zeta + iw$ represented as $Z = B\exp(i\varphi)$, where B and φ are the amplitude and the phase of Z, respectively. In these variables, the system of equations (10) is expressed as

$$\dot{B} = -\nu(B - \cos\varphi), \tag{11a}$$

$$\dot{\varphi} = u - \frac{\nu}{B}\sin\varphi, \tag{11b}$$

$$u + \theta B \sin\varphi = u_i(\tau). \tag{11c}$$

The dots in Eqs. (11) denote time derivatives. In the zero order of perturbation theory in the parameter ν, the phase φ coincides with the vector potential A ($\varphi = A$) and B is constant. When $\nu \ll 1$, B varies slowly, and, following the Volosov method [16], one can distinguish fast, $\varphi(\tau)$, and slow, $B(\tau)$, motions in Eqs. (11b) and (11c), setting B = const in these equations. Then, the equation for the phase reduces to

$$\dot{\varphi} + \bar{\theta}\sin\varphi = u_i(\tau), \tag{12}$$

where $\bar{\theta} = \theta B + \nu/B$. The general scheme of the method

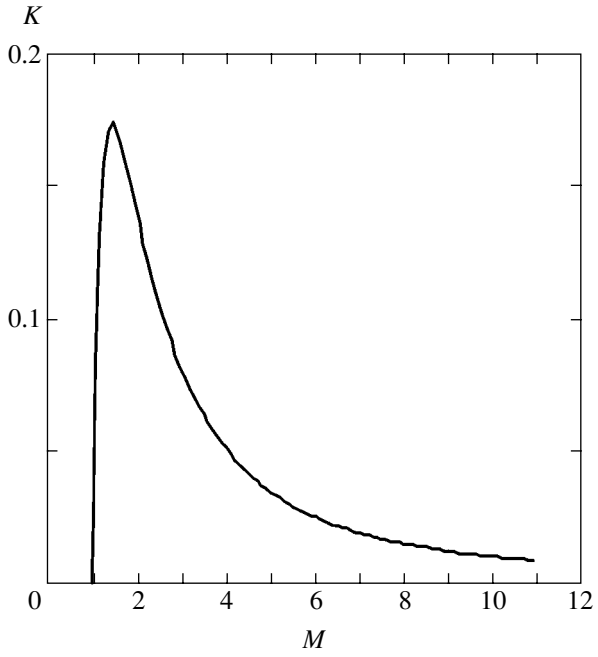

Fig. 4. Energy efficiency of generating the fundamental harmonic of a Bloch frequency as a function of the above-threshold parameter.

consists in substituting the solution to Eq. (12) for the phase, which depends on B, into Eq. (11a) and averaging the result obtained with respect to time, thus obtaining the self-consistency condition. For a stationary process ($\dot{B} = 0$), this condition is expressed as

$$B = \langle \cos\varphi \rangle, \tag{13}$$

where the angular brackets denote time averaging. Thus, when small attenuation is taken into account, Eqs. (11) reduce to an equation similar to (4) in which the coefficient multiplying the sine is renormalized. Unlike the expression $\dot{A} = u$, the relation between the intensity of the electric field in the superlattice and the phase φ becomes more general,

$$u = \frac{\theta B^2}{\nu + \theta B^2}\dot{\varphi} + \frac{\nu}{\nu + \theta B^2}u_i(\tau), \tag{14}$$

and, in the limit of $\nu \longrightarrow 0$, reduces to an ordinary relation between the field and the vector potential. As a result, one may expect that low-frequency collisions in the superlattice do not destroy the above-described excitation of Bloch oscillations and multiplication of the radiation frequency. Only the values of parameters at which these effects occur are varied. Consider, as an example, a stationary process of interaction between a lateral superlattice and a harmonic pumping field $u_i(\tau) = u_m \sin\tau$ in the limit case when $u_m \gg \theta$. Simple but rather tedious calculations yield

$$B = \langle \cos(u_m \cos\tau) \rangle = J_0(u_m),$$

where $J_0(u_m)$ are the Bessel functions. Hence, $\bar{\theta} = \theta J_0(u_m) + \nu/J_0(u_m)$. Note that $\bar{\theta} = \theta J_0(u_m) \neq 0$ for $\nu = 0$, as should be expected. Nevertheless, this fact does not contradict the theory presented above on the interaction between short electromagnetic pulses and a superlattice in the absence of attenuation since this solution corresponds to a stationary process with the relaxation time on the order of $1/\nu$. The generation of Bloch oscillations and the frequency multiplication occur only when the inequality $u_m > \bar{\theta}$ holds. The solution to this inequality and the corresponding excitation bands of Bloch oscillations are shown in Fig. 5 for $\nu = 0.1$.

Another interesting feature of the solution obtained for $\nu = 0$ is the vanishing of the parameter $\bar{\theta}$ for the amplitudes of the external field that correspond to the zeros of the Bessel function, $J_0(u_m) = 0$. This is a direct consequence of the dynamic localization of an electron in a miniband in an ac electric field [17, 18]. However, even for arbitrarily small but finite ν, the quantity $\bar{\theta}$ diverges when the Bessel function tends to zero. This suggests that, under conditions of dynamic localization, the consideration of electron collisions is essential.

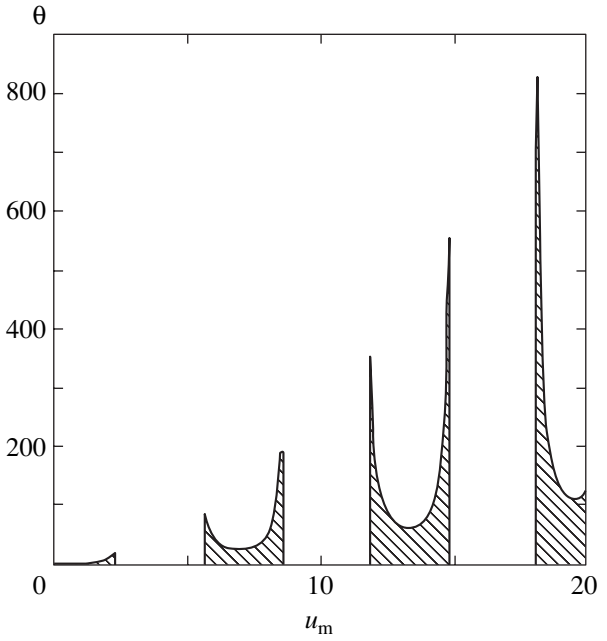

Fig. 5. Excitation bands of Bloch oscillations (dashed areas) under conditions of low-frequency collisions ($\nu = 0.1$).

In addition, we note that, when the attenuation is taken into account, the inequality for the pulse length, $\nu T_0 \ll 1$, is reversed, so that the relaxation processes of the solutions obtained cannot be considered within the approach described in this paper.

6. In conclusion, note that the possibilities considered above open up new prospects for the direct observation of Bloch oscillations and the development of highly efficient frequency multipliers for electromagnetic radiation. First, this is associated with the absence of domain instability in lateral superlattices irradiated by electromagnetic fields, which is a fundamental obstacle to the development of a Bloch oscillator based on the high-frequency instability of superlattices in dc electric fields. Second, the development of frequency multipliers based on Bloch oscillations does not require extreme conditions and may occur at room temperatures. Indeed, simple estimates show that, for a lateral superlattice of thickness $h = 100$ nm, a lateral period of $d = 20$ nm, surface concentration of carriers of $n_s = n_e h = 10^{13}$ cm^{-2}, and a parameter θ of 5, which corresponds to the carrier frequency of the incident radiation of $f \approx 0.2$ THz, the threshold value of the energy flux density of the incident wave starting from which one can observe Bloch oscillations is approximately equal to 10^4 W/cm^2. In the optimal regime, which corresponds to the maximal energy efficiency of frequency conversion, the fundamental frequency of Bloch oscillations is about 1.35 THz.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project nos. 01-02-16449 and 02-02-16385.

REFERENCES

1. S. A. Ktitorov, G. S. Simin, and V. Ya. Sindalovskiĭ, Fiz. Tverd. Tela (Leningrad) **13**, 2230 (1971) [Sov. Phys. Solid State **13**, 1872 (1971)].
2. H. Kroemer, cond-mat/0007482.
3. Yu. A. Romanov, V. P. Bovin, and L. K. Orlov, Fiz. Tekh. Poluprovodn. (Leningrad) **12**, 1665 (1978) [Sov. Phys. Semicond. **12**, 987 (1978)].
4. H. Kroemer, cond-mat/0009311.
5. D. A. Ryndyk, N. V. Demarina, J. Keller, and E. Schomburg, Phys. Rev. B **67**, 033305 (2003).
6. A. A. Ignatov, E. P. Dodin, and V. I. Shashkin, Mod. Phys. Lett. **5**, 1087 (1991).
7. A. A. Ignatov, E. P. Dodin, and A. A. Zharov, Phys. Low-Dimens. Semicond. Struct. **7**, 43 (1993).
8. L. Esaki and R. Tsu, IBM J. Res. Dev. **14**, 61 (1970).
9. E. P. Dodin, A. A. Zharov, and A. A. Ignatov, Zh. Éksp. Teor. Fiz. **114**, 2246 (1998) [JETP **87**, 1226 (1998)].
10. A. A. Zharov, E. P. Dodin, and A. S. Raspopin, Pis'ma Zh. Éksp. Teor. Fiz. **72**, 653 (2000) [JETP Lett. **72**, 453 (2000)].
11. E. P. Dodin and A. A. Zharov, Zh. Éksp. Teor. Fiz. **124**, 142 (2003) [JETP **97**, 127 (2003)].
12. H. Sakaki, Jpn. J. Appl. Phys. **28**, 314 (1989).
13. N. Noguchi, J. P. Leburton, and H. Sakaki, Inst. Phys. Conf. Ser., No. 129, 299 (1993).
14. X. L. Lei, N. J. M. Horing, and H.-L. Cui, Phys. Rev. Lett. **66**, 3277 (1991).
15. K. K. Likharev, *Introduction to the Dynamics of Josephson Junctions* (Nauka, Moscow, 1985) [in Russian].
16. V. M. Volosov, Usp. Mat. Nauk **17**, 3 (1962).
17. A. A. Ignatov and Yu. A. Romanov, Phys. Status Solidi B **73**, 327 (1976).
18. D. H. Dunlap and V. M. Kenkre, Phys. Rev. B **34**, 3625 (1986).

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 99, No. 3, 2004, pp. 559–565.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 3, 2004, pp. 642–649.

SOLIDS
Electronic Properties

Parameters of the Effective Singlet–Triplet Model for Band Structure of High-T_c Cuprates by Alternative Approaches¶

M. M. Korshunov[a,*], V. A. Gavrichkov[a], S. G. Ovchinnikov[a], Z. V. Pchelkina[b], I. A. Nekrasov[b], M. A. Korotin[b], and V. I. Anisimov[b]

[a]*Kirensky Institute of Physics, Siberian Division, Russian Academy of Sciences, Krasnoyarsk, 660036 Russia*
[b]*Institute of Metal Physics, Ural Division, Russian Academy of Sciences, Ekaterinburg, 620219 Russia*
**e-mail: mkor@iph.krasn.ru*

Received December 2, 2003

Abstract—We consider the problem of determining the parameters for high-T_c superconducting copper oxides. Alternative approaches, the *ab initio* LDA and LDA + U calculations and the generalized tight-binding (GTB) method for strongly correlated electron systems, are used to calculate hopping and exchange parameters of the effective singlet–triplet model for the CuO_2 layer. The resulting parameters are in remarkably good agreement with each other and with parameters extracted from experiment. This set of parameters is proposed for proper quantitative description of the physics of hole-doped high-T_c cuprates in the framework of effective models. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

High-T_c superconducting cuprates (HTSCs) belong to the class of substances where strong electron correlations are important. This circumstance and also the fact that these substances have nontrivial phase diagrams (see, e.g., reviews in [1]) complicate the description of HTSCs in the framework of first-principle (*ab initio*) methods, especially in the low-doping region. Therefore, the most adequate method for theoretical investigations of HTSCs is currently the model approach. Effective models of HTSCs (e.g., the *t–J* model) usually contain free parameters that can be fitted to experimental data (comparison of the calculated and experimental Fermi surfaces, dispersion curves, etc.), but the question concerning correctness of these parameters arises in the model approach. One possible way to answer this question is to obtain relations between the parameters of some effective model and the microscopic parameters of the underlying crystal structure. The underlying crystal structure of HTSCs can be described either by the three-band Emery model [2, 3] or by the multiband *p–d* model [4]. One can compare the parameters in these models with the parameters obtained by a very different approach, e.g., with *ab initio* calculated parameters. This does not mean that the *ab initio* band structure is correct. Due to strong electron correlations, it is certainly incorrect in the low-doping region, where these correlations are most significant. Nevertheless, single electron parameters are of interest and may be compared with the appropriate parameters obtained by fitting to experimental ARPES data.

In the present paper, we obtain relations between microscopic parameters of the multiband *p–d* model and parameters of the effective singlet–triplet *t–J* model for hole-doped HTSCs. We then compare these parameters and the *t–J* model parameters obtained in the *ab initio* calculations. In Section 2, the details of *ab initio* calculations within the density functional theory are presented. In Section 3, the effective singlet–triplet model is formulated as the low-energy Hamiltonian for the multiband *p–d* model with the generalized tight-binding (GTB) method applied. In both methods, the parent insulating compound La_2CuO_4 is investigated. The parameters are obtained at zero doping because, within the GTB method, the evolution of the band structure with doping is described only by changes in the occupation numbers of zero-hole, single-hole, and two-hole local terms, while all the parameters are fitted in the undoped case and are therefore fixed for all doping levels. The resulting parameters of both approaches (GTB and *ab initio*) are in very good qualitative and quantitative agreement with each other and with the parameters extracted from experiment. Also, these parameters are in reasonable agreement with the *t–J* model parameters used in the literature. We conclude that the obtained set of model parameters should be used in effective models for proper quantitative description of HTSCs in the whole doping region.

¶This article was submitted by the authors in English.

2. *AB INITIO* CALCULATION OF PARAMETERS

The band structure of La_2CuO_4 was obtained in the framework of the linear muffin-tin orbital method [5] in the tight-binding approach [6] (TB-LMTO) within the local density approximation (LDA). The crystal structure data [7] corresponds to tetragonal La_2CuO_4. The effective hopping parameters t_ρ were calculated by the least-squares fit procedure to the bands obtained in the LDA calculation [8]. The effective exchange interaction parameters J_ρ were calculated using the formula derived in [9], where the Green function method was used to calculate J_ρ as the second derivative of the ground-state energy with respect to the magnetic moment rotation angle via eigenvalues and eigenfunctions obtained in the LDA + U calculation [10]. The LDA + U approach makes it possible to obtain the experimental antiferromagnetic insulating ground state for the undoped cuprate: in contrast, the LDA approach gives a nonmagnetic metallic ground state [10]. The Coulomb parameters U = 10 eV and J = 1 eV used in the LDA + U calculation were obtained in constrained LSDA supercell calculations [11].

3. GTB METHOD AND FORMULATION OF THE EFFECTIVE SINGLET–TRIPLET MODEL

The *t–J* [12] and Hubbard [13] models are widely used to investigate HTSCs compounds. In using these models, one can in principle describe qualitatively essential physics. The parameters in these models (i.e., the hopping integral t, antiferromagnetic exchange J, and Hubbard repulsion U) are typically extracted from experimental data. Therefore, these parameters do not have a direct microscopical meaning. A more systematic approach is to write the multiband Hamiltonian for the real crystal structure (which now includes parameters of this *real* structure) and map this Hamiltonian onto some low-energy model (like the *t–J* model). In this case, parameters of the real structure could be taken from the *ab initio* calculations or fitted to experimental data.

It is convenient to use the three-band Emery *p–d* model [2, 3] or the multiband *p–d* model [4] as the starting model that properly describes crystal structure of the cuprates. The set of microscopic parameters for the first model was calculated in [14, 15]. While this model is simpler than the multiband *p–d* model, it lacks some significant features, namely, the importance of d_{z^2} orbitals on copper and p_z orbitals on apical oxygen. Nonzero occupancy of d_{z^2} orbitals was demonstrated in XAS and EELS experiments, which shows 2–10% occupancy of d_{z^2} orbitals [16, 17] and 15% doping-dependent occupancy of p_z orbitals [18] in all *p*-type (hole-doped) HTSCs. In order to take these facts into account, the multiband *p–d* model should be used,

$$H_{pd} = \sum_{f,\lambda,\sigma}(\epsilon_\lambda - \mu)n_{f\lambda\sigma} + \sum_{\langle f,g\rangle}\sum_{\lambda,\lambda',\sigma} T_{fg}^{\lambda\lambda'} c^+_{f\lambda\sigma} c_{g\lambda'\sigma} + \frac{1}{2}\sum_{f,g,\lambda,\lambda'}\sum_{\sigma_1,\sigma_2,\sigma_3,\sigma_4} V_{fg}^{\lambda\lambda'} c^+_{f\lambda\sigma_1} c_{f\lambda\sigma_3} c^+_{g\lambda'\sigma_2} c_{g\lambda'\sigma_4}, \quad (1)$$

where $c_{f\lambda\sigma}$ is the annihilation operator in the Wannier representation of the hole at site f (copper or oxygen) at orbital λ with spin σ, and $n_{f\lambda\sigma} = c^+_{f\lambda\sigma} c_{f\lambda\sigma}$. The indices λ run through $d_{x^2-y^2} \equiv d_x$ and $d_{3z^2-r^2} \equiv d_z$ orbitals on copper, p_x and p_y atomic orbitals on plane oxygen sites, and p_z orbital on apical oxygen; ϵ_λ is the single-electron energy of the atomic orbital λ; $T_{fg}^{\lambda\lambda'}$ includes hopping matrix elements between copper and oxygen (t_{pd} for hopping $d_x \leftrightarrow p_x, p_y$; $t_{pd}/\sqrt{3}$ for $d_z \leftrightarrow p_x, p_y$; t'_{pd} for $d_z \leftrightarrow p_z$) and between oxygen and oxygen (t_{pp} for hopping $p_x \leftrightarrow p_y$; t'_{pp} for hopping $p_x, p_y \leftrightarrow p_z$). The Coulomb matrix elements $V_{fg}^{\lambda\lambda'}$ include intraatomic Hubbard repulsions of two holes with opposite spins on one copper and oxygen orbital (U_d, U_p), between different orbitals of copper and oxygen (V_d, V_p), the Hund exchange on copper and oxygen (J_d, J_p), and nearest neighbor copper–oxygen Coulomb repulsion V_{pd}.

The GTB method [19] consists in exact diagonalization of the intracell part of *p–d* Hamiltonian (1) and perturbative account for the intercell part. For $La_{2-x}Sr_xCuO_4$, the unit cell is the CuO_6 cluster, and the problem of nonorthogonality of the molecular orbitals of adjacent cells is solved explicitly, by constructing the relevant Wannier functions on a five-orbital initial basis of atomic states [20, 21]. In the new symmetric basis, the intracell part of the total Hamiltonian is diagonalized, allowing one to classify all possible effective quasiparticle excitations in the CuO_2-plane according to symmetry.

Calculations [20, 21] of the quasiparticle dispersion and spectral intensities in the framework of the multiband *p–d* model using the GTB method are in very good agreement with the ARPES data on insulating compound $Sr_2CuO_2Cl_2$ [22, 23] (see Fig. 1).

Other significant results of this method are as follows [24, 25].

(i) Pinning of the Fermi level in $La_{2-x}Sr_xCuO_4$ at low concentrations was obtained in agreement with experiments [26, 27]. This pinning appears due to the in-gap state; the spectral weight of this state is proportional to the doping concentration x, and when the Fermi level reaches this in-gap band, it "stacks" there. In Fig. 2, the doping dependence of the chemical potential shift $\Delta\mu$ for *n*-type high-T_c $Nd_{2-x}Sr_xCuO_4$ (NCCO)

and *p*-type high-T_c $La_{2-x}Sr_xCuO_4$ (LSCO) is shown. The localized in-gap state also exists in NCCO for the same reason as in LSCO, but its energy is determined by the extremum of the band at the point $(\pi/2, \pi/2)$ and appears to be above the bottom of the conductivity band. Therefore, the first doped electron enters the band state at $(\pi, 0)$, and the chemical potential merges into the band for a very small concentration. At higher x, it reaches the in-gap state with pinning at $0.08 < x < 0.18$ and then μ again moves into the band. The dependence $\mu(x)$ for NCCO is quite asymmetric to the LSCO and also agrees with experimental data [26].

(ii) The experimentally observed [28] evolution of the Fermi surface with doping from the hole type (centered at (π, π)) in the underdoped region to the electron type (centered at (0, 0)) in the overdoped region is qualitatively reproduced in this method.

(iii) The pseudogap feature for $La_{2-x}Sr_xCuO_4$ is obtained as a lowering of the density of states between the in-gap state and the states at the top of the valence band.

The above results were obtained with the following set of the microscopic parameters:

$$\varepsilon_{d_{x^2-y^2}} = 0, \quad \varepsilon_{d_{3z^2-r^2}} = 0.5, \quad \varepsilon_{p_x} = 1.5,$$
$$\varepsilon_{p_z} = 0.45, \quad t_{pd} = 1, \quad t_{pp} = 0.46,$$
$$t'_{pd} = 0.58, \quad t'_{pp} = 0.42, \quad U_d = V_d = 9, \qquad (2)$$
$$J_d = 1, \quad J_p = 0, \quad U_p = V_p = 4,$$
$$V_{pd} = 1.5.$$

As the next step, we formulate the effective model. The simplest way to do this is to completely neglect the contribution of the two-particle triplet state $^3B_{1g}$. Then, there is only one low-energy two-particle state, the Zhang–Rice-type singlet $^1A_{1g}$, and the effective model is the usual *t*–*J* model. However, in the multiband *p*–*d* model, the difference $\epsilon_T - \epsilon_S$ between the energies of the two-particle singlet and the two-particle triplet depends strongly on various model parameters, particularly on the distance of apical oxygen from planar oxygen, the energy of apical oxygen, and the difference between the $d_{3z^2-r^2}$- and $d_{x^2-y^2}$-orbital energies. For realistic values of the model parameters, $\varepsilon_T - \varepsilon_S$ is close to 0.5 eV [21, 32], in contrast to the three-band model, where this value is about 2 eV (this case was considered in [29, 30]). To take the triplet states into account, we derive the effective Hamiltonian for the multiband *p*–*d* model by exclusion of the intersubband hopping between lower (LHB) and upper (UHB) Hubbard subbands, similarly to [12].

The Hubbard *X*-operator $X_f^{pq} \equiv |p\rangle\langle q|$ on site *f* represents a natural language to describe strongly correlated electron systems, and we therefore use these operators in the rest of the paper. The *X*-operators are constructed in the Hilbert space, which consists of the vacuum $n_h = 0$ state $|0\rangle$, the single-hole $|\sigma\rangle = \{|\uparrow\rangle, |\downarrow\rangle\}$ state of b_{1g} sym-

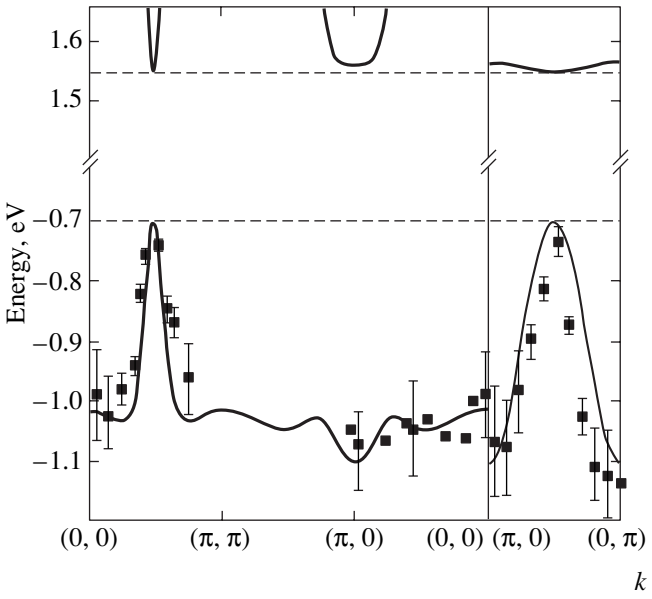

Fig. 1. The GTB method dispersion (doping concentration $x = 0$) of the top of the valence band and the bottom of the conduction band divided by the insulating gap. Horizontal dashed lines mark the in-gap states whose spectral weight is proportional to x. Points with error bars represent experimental ARPES data for $Sr_2CuO_2Cl_2$ [22].

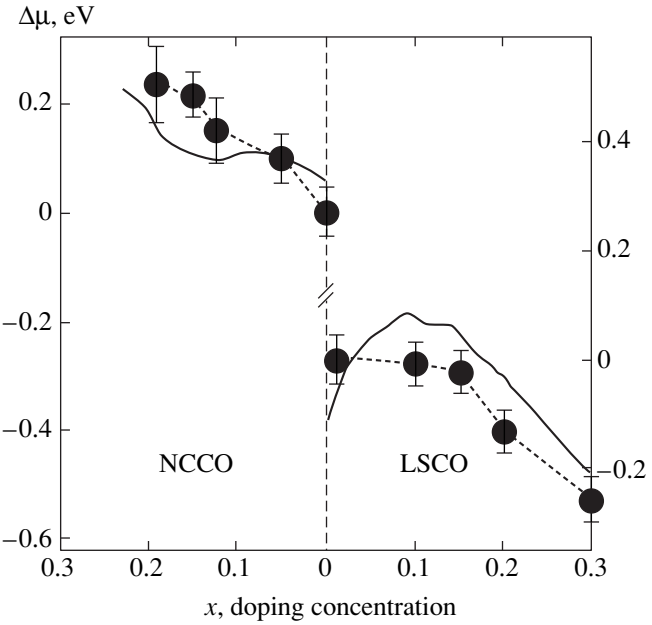

Fig. 2. Dependence of the chemical potential shift $\Delta\mu$ on the concentration of doping x for $Nd_{2-x}Sr_xCuO_4$ and $La_{2-x}Sr_xCuO_4$. Straight lines are results of the GTB calculations; filled circles with error bars are experimental points [26].

Table 1. Parameters of the effective singlet–triplet model for *p*-type cuprates obtained in the framework of the GTB method (all values in eV)

ρ	t_ρ^{00}	t_ρ^{SS}	t_ρ^{0S}	t_ρ^{TT}	t_ρ^{ST}	J_ρ
(0, 1)	0.373	0.587	−0.479	0.034	0.156	0.115
(1, 1)	0.002	−0.050	0.015	−0.011	0	0.0001
(0, 2)	0.050	0.090	−0.068	0.015	0.033	0.0023
(2, 1)	0.007	0.001	−0.006	−0.004	0.001	0

metry, the two-hole singlet state $|S\rangle$ of ${}^1A_{1g}$ symmetry, and the two-hole triplet state $|TM\rangle$ (where $M = +1, 0, -1$) of ${}^3B_{1g}$ symmetry.

We write the Hamiltonian as $H = H_0 + H_1$, where the excitations via the charge transfer gap E_{ct} are included in H_1. We then define the operator $H(\epsilon) = H_0 + \epsilon H_1$ and perform the unitary transformation $\tilde{H}(\epsilon) = \exp(-i\epsilon\hat{S})H(\epsilon)\exp(i\epsilon\hat{S})$. The vanishing of the term linear in ϵ in $\tilde{H}(\epsilon)$ gives the equation for the matrix $\hat{S}$, $H_1 + i[H_0, \hat{S}] = 0$. The effective Hamiltonian is obtained in the second order in ϵ; at $\epsilon = 1$, it is given by

$$\tilde{H} = H_0 + \frac{1}{2}i[H_1, \hat{S}]. \tag{3}$$

Thus, for the multiband *p–d* model (1) in the case of electron doping (*n*-type systems), we obtain the usual *t–J* model,

$$\begin{aligned} H_{t-J} &= \sum_{f,\sigma}\varepsilon_1 X_f^{\sigma\sigma} + \sum_{\langle f,g\rangle,\sigma} t_{fg}^{00} X_f^{\sigma 0} X_g^{0\sigma} \\ &+ \sum_{\langle f,g\rangle} J_{fg}\left(\mathbf{S}_f\cdot\mathbf{S}_g - \frac{1}{4}n_f n_g\right), \end{aligned} \tag{4}$$

where $\mathbf{S}_f$ are spin operators and n_f are the particle number operators. The term $J_{fg} = 2(t_{fg}^{0S})^2/E_{ct}$ is the exchange integral, and E_{ct} is the energy of the charge-transfer gap (similar to U in the Hubbard model, $E_{ct} \approx 2$ eV for cuprates). The chemical potential μ is included in ε_1.

For *p*-type systems, the effective Hamiltonian has the form of a singlet–triplet *t–J* model [31],

$$H = H_0 + H_t + \sum_{\langle f,g\rangle} J_{fg}\left(\mathbf{S}_f\cdot\mathbf{S}_g - \frac{1}{4}n_f n_g\right), \tag{5}$$

where H_0 (the unperturbed part of the Hamiltonian) and H_t (the kinetic part of H) are given by

$$H_0 = \sum_f\left[\varepsilon_1\sum_\sigma X_f^{\sigma\sigma} + \varepsilon_{2S}X_f^{SS} + \varepsilon_{2T}\sum_M X_f^{TMTM}\right],$$

$$\begin{aligned} H_t = \sum_{\langle f,g\rangle,\sigma}\{ & t_{fg}^{SS}X_f^{S\bar{\sigma}}X_g^{\bar{\sigma}S} \\ & + t_{fg}^{TT}(\sigma\sqrt{2}X_f^{T0\bar{\sigma}} - X_f^{T2\sigma\sigma})(\sigma\sqrt{2}X_g^{\bar{\sigma}T0} - X_g^{\sigma T2\sigma}) \\ & + t_{fg}^{ST}2\sigma\gamma_b[X_f^{S\bar{\sigma}}(\sigma\sqrt{2}X_g^{\bar{\sigma}T0} - X_g^{\sigma T2\sigma}) + \text{H.c.}]\}. \end{aligned}$$

The superscripts of hopping integrals (0, *S*, *T*) correspond to excitations that are accompanied by hopping from site *f* to *g*, i.e., the Hamiltonian involves the terms $\sum_{\langle f,g\rangle,\sigma} t_{fg}^{MN}X_f^{M\sigma}X_g^{\sigma N}$. The relation between these effective hoppings and microscopic parameters of the multiband *p–d* model is as follows:

$$\begin{aligned} t_{fg}^{00} &= -2t_{pd}\mu_{fg}2uv - 2t_{pp}\nu_{fg}v^2, \\ t_{fg}^{SS} &= -2t_{pd}\mu_{fg}2\gamma_x\gamma_b - 2t_{pp}\nu_{fg}\gamma_b^2, \\ t_{fg}^{0S} &= -2t_{pd}\mu_{fg}(v\gamma_x + u\gamma_b) - 2t_{pp}\nu_{fg}v\gamma_b, \\ t_{fg}^{TT} &= \frac{2t_{pd}}{\sqrt{3}}\lambda_{fg}2\gamma_a\gamma_z + 2t_{pp}\nu_{fg}\gamma_a^2 - 2t'_{pp}\lambda_{fg}2\gamma_p\gamma_a, \\ t_{fg}^{ST} &= \frac{2t_{pd}}{\sqrt{3}}\xi_{fg}\gamma_z + 2t_{pp}\chi_{fg}\gamma_a - 2t'_{pp}\xi_{fg}\gamma_p. \end{aligned} \tag{6}$$

The factors μ, ν, λ, ξ, χ are the coefficients of the Wannier transformation performed in the GTB method and u, v, γ_a, γ_b, γ_z, γ_p are the matrix elements of the annihilation and creation operators in the Hubbard *X*-operator representation.

The resulting Hamiltonian (5) is the generalization of the *t–J* model to account for the two-particle triplet state. A significant feature of the effective singlet–triplet model is the asymmetry of *n*- and *p*-type systems, which is known experimentally. We can therefore conclude that, for *n*-type systems, the usual *t–J* model is applicable, while for *p*-type superconductors with complicated structure at the top of the valence band, the singlet–triplet transitions play an important role.

Using the set of microscopic parameters (2) in Table 1, we present numerical values of the hopping and exchange parameters calculated in accordance with (6).

4. COMPARISON OF PARAMETERS

The resulting parameters from *ab initio* [8] and GTB calculations are presented in Table 2. Here, ρ is the connecting vector between two copper centers, t_ρ is the

Table 2. Comparison of *ab initio* parameters [8] and parameters obtained in the framework of the GTB method (all values in eV)

	ab initio		GTB method	
ρ	t_ρ	J_ρ	t_ρ	J_ρ
(0, 1)	0.486	0.109	0.587	0.115
(1, 1)	−0.086	0.016	−0.050	0.0001
(0, 2)	−0.006	0	0.090	0.0023
(2, 1)	0	0	0.001	0

hopping parameter (equal to t_ρ^{SS}, see (5) and (6), in the effective singlet–triplet model), and J_ρ is the antiferromagnetic exchange integral.

As one can see, despite slight differences, the parameters in both methods are very close and show similar dependence on distance. It is worth mentioning that both methods give a disproportionality between t_ρ^2 and J_ρ. In the usual *t–J* model, the proportionality $J_\rho = 2t_\rho^2/U$ occurs as soon as this *t–J* model is obtained from the Hubbard model with the Hubbard repulsion *U*. In the singlet–triplet model, the intersubband hopping t_ρ^{0S} that determines the value of J_ρ is different from the intrasubband hopping t_ρ^{SS} that determines t_ρ. This leads to a more complicated relation between t_ρ and J_ρ.

In the framework of the LDA band structure of YBa_2CuO_{7+x} and within the orbital projection approach, it was shown [33] that the one-band Hamiltonian reduced from the eight-band Hamiltonian should include not only the nearest neighbor hopping terms (*t*), but also second (*t*') and third (*t*") nearest neighbor hoppings. In the GTB method, the dependence of the hoppings t_ρ on distance automatically results from the distance dependence of the coefficients of the Wannier transformation performed in this method (see Eq. (6)). To show the correspondence between the results of different authors, we compare our parameters and the parameters widely used by different groups in Table 3.

The parameters extracted from experimental data are listed in columns I–VI of Table 3. The LDA calculated parameters are presented in columns VII and VIII. Our results for hoppings agree best with columns III, VII, and VIII. This similarity is not surprising. In the LDA calculations, the bandwidth of strongly correlated electron systems is usually overestimated because the strong Coulomb repulsion of electrons is not taken into account properly. However, it is well known that the Fermi surface obtained by this method is in very good agreement with experiments. The main contribution to the shape of the Fermi surface comes from the kinetic energy of the electrons (hopping parameters), and therefore the values of hoppings should be properly estimated by the LDA calculations (columns VII, VIII). In [37, 38] (column III), the parameters were obtained by fitting the LSCO tight-binding Fermi surface to the experimental one. This procedure gives the same values as the LDA calculation. By the same technique, the parameters for $Bi_2Sr_2CaCu_2O_{8+x}$ (Bi2212, column IV) were obtained [37, 38]. These parameters are different from those in the LSCO case and in the present paper; the most straightforward explanation is a more complicated structure of the Fermi surface of compound Bi2212. In the present paper, single-layer (LSCO-like) compounds are considered and the effects of multiple CuO_2 planes (i.e., bilayer splitting) are neglected. The difference between our hoppings and hoppings in column V

Table 3. Comparison of the calculated parameters and parameters used in the literature

Quantity	0*	0**	I***	II***	III***	IV***	V***	VI***	VII****	VIII****	IX*****	X*****
	LSCO	LSCO	LSCO	LSCO	LSCO	Bi2212	YBCO	SCOC	YBCO	LSCO	LSCO	YBCO
	this work	this work	[34]	[35, 36]	[37, 38]	SCOC [37, 38, 39]	[40]	[41]	[33]	[42]	[43]	[43]
t, eV	0.587	0.486	0.416	0.35	0.35	0.35	0.40	0.40	0.349	0.43	–	–
t'/*t*	−0.085	−0.18	−0.350	−0.20	−0.12	−0.34	−0.42	−0.35	−0.028	−0.17	–	–
t"/*t*	0.154	0.012	–	0.15	0.08	0.23	−0.25	0.25	0.178	–	–	–
J, eV	0.115	0.109	0.125	0.14	0.14	0.14	0.17	0.12	–	–	0.126	0.125, 0.150
J/|*t*|	0.196	0.224	0.300	0.40	0.40	0.40	0.43	0.30	–	–	–	–

* GTB method parameters.

** *Ab initio* parameters obtained in the present paper.

*** Parameters obtained by fitting to experimental data.

**** *Ab initio* parameters.

***** Parameters obtained from two-magnon Raman scattering.

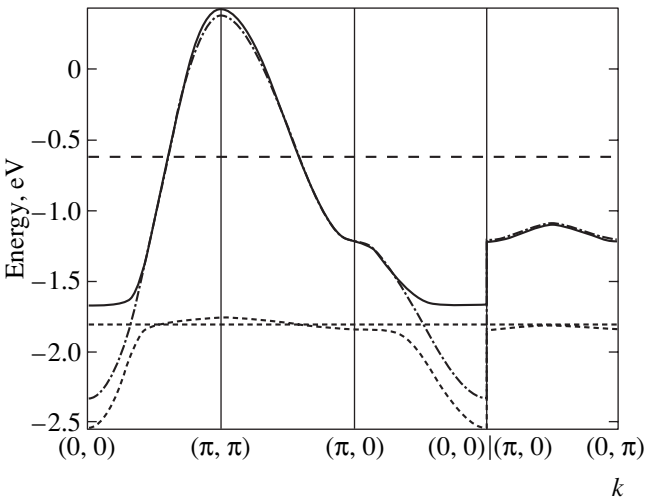

Fig. 3. Dispersion curves on top of the valence band for the effective singlet–triplet model (singlet subband is shown with solid line, triplet subbands with dotted lines) and the *t*–*t*'–*J* model (dash-dotted line) at the optimal doping $x = 0.15$; the dashed line represents the self-consistently obtained chemical potential μ.

appears due to the same reason (in [40], the $YBa_2Cu_3O_6$ insulating compound was investigated).

In the last two columns of Table 3, the antiferromagnetic exchange parameters *J* obtained from the two-magnon Raman scattering analysis by momentum expansion (LSCO, column IX) and spin-wave theory (YBCO, column X) are presented (for details, see review [43] and references therein). Our values of *J* (column 0) are in good agreement with the values extracted from experiments and similar to those listed in columns I–VI.

In [44], the Heisenberg Hamiltonian on the square lattice with plaquette ring exchange was investigated. The fitted exchange interactions $J = 0.151$ eV, $J' = J'' = 0.025J$ give the values for the spin stiffness and the Néel temperature in excellent agreement with experimental data for insulating compound La_2CuO_4. In the GTB calculations, $J = 0.115$ eV, $J' = 0.0009J$, and $J'' = 0.034J$. The values of *J* are close to each other but different. This difference could be explained by the fact that authors of [44] used the Heisenberg Hamiltonian and inclusion of the hopping term should renormalize the presented exchange interaction values. Agreement between *J*" in the GTB calculations and in [44] is good but the values of *J*' are completely different. The latter could be attributed to oversimplification of calculations in [44], where the authors set $J' = J''$ by hand to restrict the number of fitting parameters.

We now discuss the difference between our parameters and the parameters in columns I, II, VI, and column IV (SCOC). The hoppings in the papers cited above were obtained by fitting the *t*–*t*'–*t*"–*J* model dispersion to the experimental ARPES spectra [22, 39] for insulating $Sr_2CuO_2Cl_2$. We claim that the discrepancy between the GTB method results and the *t*–*t*'–*t*"–*J* model results stems from the absence of singlet–triplet hybridization in the latter model. This statement can be proved by comparing the dispersion in the "bare" *t*–*t*'–*J* model (4) and in the singlet–triplet *t*–*t*'–*J* model (5). The paramagnetic nonsuperconducting phase was investigated in the Hubbard-I approximation in both the singlet–triplet and *t*–*t*'–*J* models. The results for optimal doping (with the concentration of holes $x = 0.15$) are presented in Fig. 3.

There is a strong mixing of singlet and triplet bands along the (0, 0)–(π, π) and (π, 0)–(0, 0) directions due to the t^{ST} matrix element (see (6)) in both paramagnetic (Fig. 3) and antiferromagnetic phases (Fig. 1). It is exactly the admixture of the triplet states that determines coincidence of the dispersion in our approach and the ARPES data in the undoped SCOC at the energies 0.3–0.4 eV below the top of the valence band, where the *t*–*t*'–*J* model [34] fails and the *t*–*t*'–*t*"–*J* model involves the additional parameter *t*" [35, 37]. In our approach, this parameter is not as necessary as in the "bare" *t*–*t*'–*J* model, because the singlet–triplet hybridization is included explicitly.

In [45], the *t*–*t*'–*t*"–*J* model was also used to describe the dispersion of insulating $Sr_2CuO_2Cl_2$, with the same set of parameters as in [37, 38]. However, the authors of [45] used a totally different definition of hopping parameters: in their paper, the *t*' stands for hopping between two nearest neighbor oxygens and the *t*" stands for the hopping between two oxygens on the two sides of Cu. Such a definition is completely different from that used in other papers cited, where *t*, *t*', *t*" stand for hoppings between plaquettes centered on copper sides, and we cannot therefore compare our results with theirs.

The analysis of the data in Table 3 gives the following ranges for parameters: 0.350–0.587 for *t*, (–0.420)–(–0.028) for *t*'/*t*, 0.012–0.250 for *t*"/*t* with the exception of the value in [40], and 0.115–0.150 eV for *J*. In general, we see a close similarity in the first-neighbor hopping *t* and the interaction *J* for the different methods and materials, and greater discrepancy in subtle parameters such as *t*' and *t*".

5. CONCLUSIONS

One of the significant results in this paper is the relation (6) between microscopic parameters and parameters of the effective singlet–triplet model. The effective model parameters are therefore no longer free and have a direct physical meaning coming from the dependence on microscopic parameters. The parameters of the effective singlet–triplet model were obtained from both *ab initio* and model calculations. Model calculations were performed in the framework of the GTB method for insulating single-layer copper oxide superconductor. The *ab initio* calculations for La_2CuO_4 were performed by the conventional LDA TB-LMTO method.

The agreement between the parameters is remarkably good. The parameters obtained also agree well with widely used parameters of the t–t'–t''–J model, although a certain difference exists. This difference is attributed to disregard of triplet excitations in the simple t–t'–t''–J model. After careful analysis, we proposed the set of parameters for effective models (e.g., the t–t'–t''–J model or the effective singlet–triplet model) for proper quantitative description of physics of hole-doped high-T_c cuprates.

ACKNOWLEDGMENTS

M.M.K., V.A.G., and S.G.O. thank the Free University of Berlin for hospitality during their stay. This work was supported by INTAS (grant no. 01-0654), the Joint Integration Program of the Siberian and Ural Divisions of the Russian Academy of Sciences, the Russian Foundation for Basic Research (project nos. 03-02-16124, 04-02-16096, and 03-02-39024), the Russian Federal Program "Integration" (grant no. B0017), the Program of the Physics Division of the Russian Academy of Sciences "Strongly Correlated Electron Systems," the Siberian Division of the Russian Academy of Sciences (Lavrent'ev Contest for Young Scientists), Dynasty Foundation and International Center for Fundamental Physics in Moscow (I.A.N., M.M.K.), and Russian Science Support Foundation Program for Young PhD of Russian Academy of Sciences 2004 (I.A.N.).

REFERENCES

1. Z.-X. Shen and D. S. Dessau, Phys. Rep. **253**, 1 (1995); E. Dagotto, Rev. Mod. Phys. **66**, 763 (1994); A. P. Kampf, Phys. Rep. **249**, 219 (1994).
2. V. J. Emery, Phys. Rev. Lett. **58**, 2794 (1987).
3. C. M. Varma, S. Schmitt-Rink, and E. Abrahams, Solid State Commun. **62**, 681 (1987).
4. Yu. Gaididei and V. Loktev, Phys. Status Solidi B **147**, 307 (1988).
5. O. K. Andersen and O. Jepsen, Phys. Rev. Lett. **53**, 2571 (1984).
6. O. K. Andersen, Z. Pawlowska, and O. Jepsen, Phys. Rev. B **34**, 5253 (1986).
7. J. D. Axe and M. K. Crawford, J. Low Temp. Phys. **95**, 271 (1994).
8. V. I. Anisimov, M. A. Korotin, I. A. Nekrasov, *et al.*, Phys. Rev. B **66**, 100502 (2002).
9. A. I. Lichtenstein *et al.*, J. Magn. Magn. Mater. **67**, 65 (1987); A. I. Lichtenstein, V. I. Anisimov, and J. Zaanen, Phys. Rev. B **52**, R5467 (1995).
10. V. I. Anisimov, J. Zaanen, and O. Andersen, Phys. Rev. B **44**, 943 (1991); V. I. Anisimov *et al.*, J. Phys.: Condens. Matter **9**, 767 (1997).
11. O. Gunnarsson *et al.*, Phys. Rev. B **39**, 1708 (1989); V. I. Anisimov and O. Gunnarsson, Phys. Rev. B **43**, 7570 (1991).
12. K. A. Chao, J. Spalek, and A. M. Oles, J. Phys. C: Solid State Phys. **10**, 271 (1977).
13. J. C. Hubbard, Proc. R. Soc. London, Ser. A **276**, 238 (1963).
14. M. S. Hybertsen, M. Schluter, and N. E. Christensen, Phys. Rev. B **39**, 9028 (1989).
15. A. K. McMahan, J. F. Annett, and R. M. Martin, Phys. Rev. B **42**, 6268 (1990).
16. A. Bianconi *et al.*, Phys. Rev. B **38**, 7196 (1988).
17. H. Romberg *et al.*, Phys. Rev. B **41**, 2609 (1990).
18. C. H. Chen *et al.*, Phys. Rev. Lett. **68**, 2543 (1992).
19. S. G. Ovchinnicov and I. S. Sandalov, Physica C (Amsterdam) **161**, 607 (1989).
20. V. A. Gavrichkov, A. A. Borisov, and S. G. Ovchinnikov, Phys. Rev. B **64**, 235124 (2001).
21. V. A. Gavrichkov, S. G. Ovchinnikov, A. A. Borisov, and E. G. Goryachev, Zh. Éksp. Teor. Fiz. **118**, 422 (2000) [JETP **91**, 369 (2000)].
22. B. O. Wells *et al.*, Phys. Rev. Lett. **74**, 964 (1995).
23. C. Dürr *et al.*, Phys. Rev. B **63**, 014505 (2001).
24. A. A. Borisov, V. A. Gavrichkov, and S. G. Ovchinnikov, Mod. Phys. Lett. B **17**, 479 (2003).
25. A. A. Borisov, V. A. Gavrichkov, and S. G. Ovchinnikov, Zh. Éksp. Teor. Fiz. **124**, 862 (2003) [JETP **97**, 773 (2003)].
26. N. Harima *et al.*, Phys. Rev. B **64**, 220507(R) (2001).
27. A. Ino *et al.*, Phys. Rev. Lett. **79**, 2101 (1997).
28. A. Ino *et al.*, Phys. Rev. B **65**, 094504 (2002).
29. J. Zaanen, A. M. Oles, and P. Horsch, Phys. Rev. B **46**, 5798 (1992).
30. R. Hayn *et al.*, Phys. Rev. B **47**, 5253 (1993).
31. M. Korshunov and S. Ovchinnikov, Fiz. Tverd. Tela (St. Petersburg) **43**, 399 (2001) [Phys. Solid State **43**, 416 (2001)].
32. R. Raimondi *et al.*, Phys. Rev. B **53**, 8774 (1996).
33. O. K. Andersen *et al.*, J. Phys. Chem. Solids **56**, 1573 (1995).
34. A. Nazarenko *et al.*, Phys. Rev. B **51**, 8676 (1995).
35. V. I. Belinicher, A. I. Chernyshev, and V. A. Shubin, Phys. Rev. B **53**, 335 (1996).
36. V. I. Belinicher, A. I. Chernyshev, and V. A. Shubin, Phys. Rev. B **54**, 14914 (1996).
37. T. Tohayama and S. Maekawa, Supercond. Sci. Technol. **13**, R17 (2000).
38. T. Tohayama and S. Maekawa, Phys. Rev. B **67**, 092509 (2003).
39. C. Kim *et al.*, Phys. Rev. Lett. **80**, 4245 (1998).
40. F. P. Onufrieva, V. P. Kushnir, and B. P. Toperverg, Phys. Rev. B **50**, 12935 (1994).
41. R. Eder, Y. Ohta, and G. A. Sawatzky, Phys. Rev. B **55**, R3414 (1997).
42. E. Pavarini *et al.*, Phys. Rev. Lett. **87**, 047003 (2001).
43. W. Brenig, Phys. Rep. **251**, 153 (1995).
44. A. A. Katanin and A. P. Kampf, Phys. Rev. B **66**, 100403(R) (2002).
45. T. Xiang and J. M. Wheatley, Phys. Rev. B **54**, R12653 (1996).

Journal of Experimental and Theoretical Physics, Vol. 99, No. 3, 2004, pp. 566–573.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 3, 2004, pp. 650–658.

SOLIDS
Electronic Properties

The Mechanism of Suppression of Strong Electron Correlations in $FeBO_3$ at High Pressures

A. G. Gavriliuk[a,b], I. A. Trojan[b], S. G. Ovchinnikov[c], I. S. Lyubutin[a], and V. A. Sarkisyan[a]

[a]*Shubnikov Institute of Crystallography, Russian Academy of Sciences, Leninskiĭ pr. 59, Moscow, 119333 Russia*
[b]*Institute of High-Pressure Physics, Russian Academy of Sciences, Troitsk, Moscow region, 142092 Russia*
[c]*Kirensky Institute of Physics, Siberian Division, Russian Academy of Sciences, Krasnoyarsk, 660036 Russia*
e-mail: lyubutin@ns.crys.ras.ru
Received December 9, 2003

Abstract—The optical absorption spectra of iron borate ($FeBO_3$) are measured at high pressures up to P = 82 GPa. A mechanism of suppression of strong electron correlations is proposed within the framework of the generalized tight binding method, which leads to the experimentally observed magnetic, electronic, and structural phase transitions. Taking into account peculiarities of the crystal structure of $FeBO_3$ and the strong s–p hybridization of boron and oxygen, it is established that, as the distance between ions varies with increasing pressure, the crystal field parameter begins to play a decisive role in the electron transitions, while the influence of the d band broadening is negligibly small. Parameters of the theory are calculated as functions of the pressure. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Recently [1–4], we have reported on a number of phase transitions induced by high pressure in iron borate ($FeBO_3$). In particular, it was established that a transition of the magnet–nonmagnet type with collapse of the localized magnetic moment takes place at a pressure of about 47 GPa [1, 2]. An insulator–semiconductor type transition is observed at approximately the same pressure [3], and a structural phase transition with symmetry conservation ($R\bar{3}c \longrightarrow R\bar{3}c$) and a 9% jump in the unit cell volume takes place at about 53 GPa [4]. A model describing the electron structure of $FeBO_3$ and its variation at high pressures [5] was developed within the framework of the generalized tight binding method for systems with strong electron correlations. However, several experimentally observed effects and some principal questions concerning the physical nature of phase transitions still remain unclear. These questions are related to factors leading to the magnetic collapse and to the fact that the electron system exhibits the transition to a semiconducting state rather than to the metallic state. It is also unclear why the magnetic and electronic transitions occur at the same pressure, while the structural transition takes place at a different pressure.

In this paper, we report on the results of experimental investigation of the optical absorption spectra of single crystal $FeBO_3$ in a range of pressures up to 82 GPa. The dependences of the absorption spectra and the electron structure of $FeBO_3$ on the pressure have been theoretically studied in both low- and high-pressure phases. The theory employs the generalized tight binding method for a multiband p–d model [5] and combines the concepts of quasi-particles from the Landau theory of a Fermi liquid with Hubbard's notions of the predominant role of intraatomic Coulomb interactions in systems with strong electron correlations. Parameters of the theory calculated using the experimental data are consistent with the results of observations. The proposed theory qualitatively explains many of the effects experimentally observed in $FeBO_3$ crystals—in particular, the transition to a semiconducting (rather than metallic) state—and predicts a transition to the metallic state with further increase in pressure.

2. EXPERIMENTAL

High-quality, optically transparent $FeBO_3$ single crystals of light-green color were grown from by the flux method. The optical absorption spectra were measured at room temperature in a range of pressures up to 82 GPa in a cell with diamond anvils. The diamond anvil culets were about 400 μm in diameter, and a hole at the center of a rhenium gasket had a diameter of ~120 μm. The measurements were performed on ~50 × 50 μm^2 plates of various thicknesses from ~2 to ~15 μm. The plates were cut from a massive $FeBO_3$ single crystal so that their large faces coincided with the (111) basal plane.

1063-7761/04/9903-0566$26.00 © 2004 MAIK "Nauka/Interperiodica"

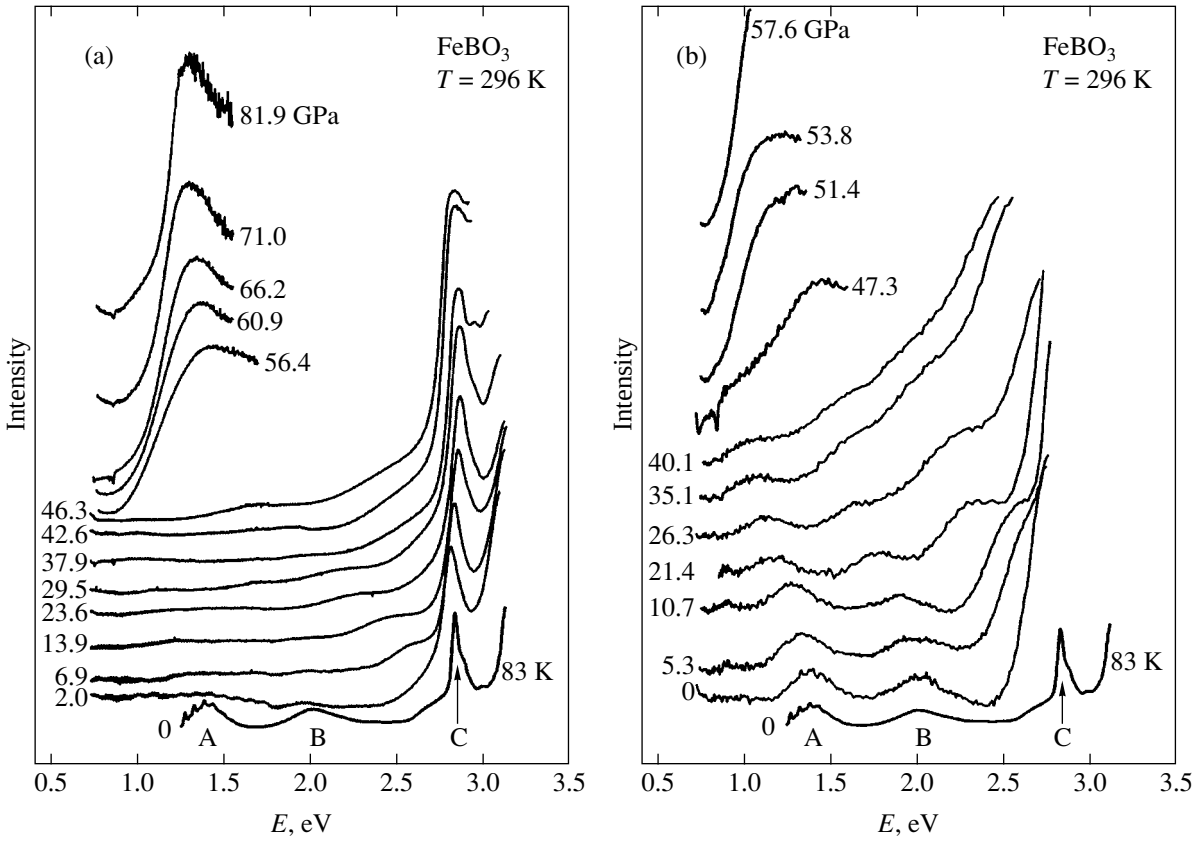

Fig. 1. Evolution of the room-temperature absorption spectra (A, B, C are absorption bands) measured at various pressures for (a) thin (d = 2–3 μm) and (b) thick (d = 10–15 μm) single crystal $FeBO_3$ samples (figures at the curves indicate pressure in GPa; the bottom curve shows the spectrum measured at 83 K [6]).

During the spectral measurements, the light beam in the high-pressure cell was directed perpendicular to the basal plane of the crystal and focused into a 20-μm-diameter spot on the sample surface. Polyethyl siloxane (PES-5) was used as a pressure-transmitting medium providing quasi-hydrostatic conditions. After pressure release, the single crystal samples remained intact.

The absorption spectra at high pressure were measured in the visible and near-IR regions in the wavelength range from 0.3 to 5 μm. The optical setup and the experimental procedure were described in detail elsewhere [3]. The absorption spectrum was calculated by the standard method using the formula

$$I = I_0 \exp(-\alpha d),$$

where I_0 is the reference beam intensity, d is the sample thickness, and α is the optical absorption coefficient.

3. EXPERIMENTAL RESULTS: EFFECT OF PRESSURE ON THE POSITIONS OF OPTICAL ABSORPTION BANDS

Figure 1a shows evolution of the optical absorption spectra of single crystal $FeBO_3$ with increasing pressure P at room temperature. These measurements were performed on samples with a thickness of 2–3 μm. As the pressure grows to 46 GPa, the energy of the optical absorption edge slowly increases. At $P \approx 46$ GPa, the absorption edge energy drops abruptly from ~3 to ~0.8 eV and then virtually does not change as the pressure is further increased up to 82 GPa (Fig. 2). The jump in the optical gap at 46 GPa correlates with the magnetic moment collapse observed previously [1, 2] at the same pressure.

The room-temperature spectra of "thick" samples (d = 10–15 μm) also clearly reveal the evolution of the absorption bands related to various excitation processes in the electron system of $FeBO_3$ (Fig. 1b). The positions of maxima of the absorption bands denoted by capital letters A, B and C at a zero applied pressure correspond to the energies 1.395 ± 0.006, 2.029 ± 0.022, and 2.803 ± 0.005 eV, respectively. For the comparison, Fig. 1 also shows the absorption spectrum of $FeBO_3$ measured at 83 K [6].

According to our experimental results, band C exhibits splitting into two signals (C and C1) with increasing pressure. The absorption band C has the shape of a narrow peak (Fig. 1a) and its position can be

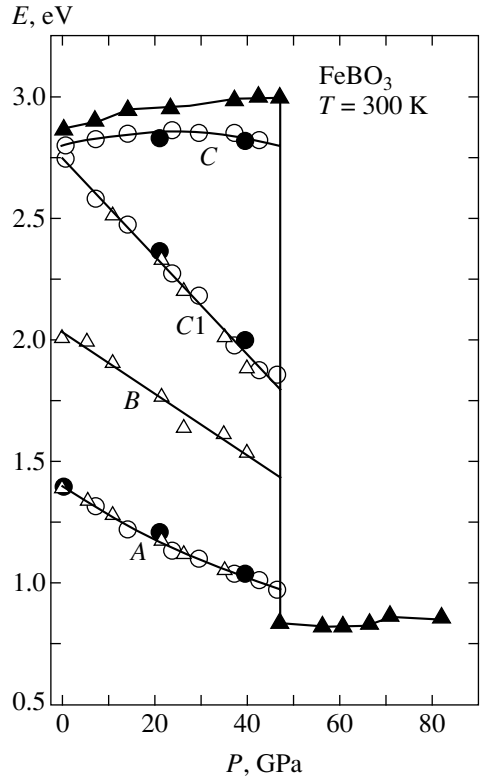

Fig. 2. Plots of the energy positions of the optical absorption edge and absorption peaks A, B, C, and C1 versus pressure for single crystal $FeBO_3$ measured at room temperature in several series of experiments (different symbols refer to samples of various thicknesses).

determined with a much better accuracy than positions of the other absorption bands. The energy of peak C weakly varies with increasing pressure as described by a quadratic law and exhibits a maximum in the region of 25 GPa.

Figure 2 shows the pressure dependence of the positions of various absorption bands and the optical absorption edge. Parameters of the electron transitions corresponding to these bands are given in the table. These parameters and related errors were calculated from experimental data using a linear approximation. For bands A and C, the experimental data were approximated using a second-degree polynomial.

4. ELECTRON STRUCTURE OF $FeBO_3$ AND ITS PRESSURE DEPENDENCE IN MANY-ELECTRON MODEL

In the absence of applied pressure, the optical absorption spectrum of $FeBO_3$ exhibits three main bands (A, B, and C) occurring below 2.9 eV inside the bandgap. According to the traditional interpretation, these bands are due to the following exciton transitions in the d^5 configuration [7–10]:

$$\omega_A = E(^4T_1) - E(^6A_1),$$
$$\omega_B = E(^4T_2) - E(^6A_1), \quad (1)$$
$$\omega_C = E(^4A_1) - E(^6A_1).$$

However, the temperature dependences of the energies of bands A, B, and C, as well as the circular dichroism spectra [11, 12], are significantly different. This poses questions concerning adequacy of the aforementioned interpretation. In particular, the intensity of peak C is much greater than the intensities of bands A and B, which suggests that an additional allowed optical transition may exist in the vicinity of ω_C. Such a transition can be related to the electron excitation with charge transfer $p^6d^5 \longrightarrow p^5d^6$. Both *d–d* excitations from the ground state to the upper d^5 terms and the $d^5 \longrightarrow d^6$ transitions in many-electron d^n configurations take place under conditions of strong electron correlations. These effects cannot be adequately described using one-electron methods of the band theory.

Recently [5], a many-electron model of the electron structure of $FeBO_3$ has been developed with allowance for strong electron correlations. The absorption spectrum calculated within the framework of this model exhibits additional excitations with charge transfer [6]. The model has been described in detail in [6], but the effect of pressure on the electron structure was not considered. For this reason, we present a brief outline of this model and consider the pressure-induced changes in the electron structure of $FeBO_3$.

The model is based on *ab initio* one-electron band calculations of the electron structure of $FeBO_3$ using the density functional method in the local spin density approximation [13] and on calculations of the molecu-

The energies E_0 of optical transitions at zero applied pressure and their baric derivatives at high pressures in single crystal $FeBO_3$ at room temperature. The last column presents theoretical *dE*/*dP* values (see Section 5)

Optical transition	E_0, eV	dE/dP, eV/GPa	d^2E/dP^2, eV/(GPa)^2	$(dE/dP)_{theor}$, eV/GPa
A	1.395 ± 0.006	-0.0115 ± 0.0007	$(5.69 \pm 1.56) \times 10^{-5}$	−0.0156
B	2.029 ± 0.022	-0.0125 ± 0.0009	–	−0.0158
C	2.803 ± 0.005	$+0.0051 \pm 0.0006$	$(-1.02 \pm 0.13) \times 10^{-4}$	0
C1	2.749 ± 0.017	-0.0199 ± 0.0006	–	−0.0174

lar orbitals of a FeB_6O_6 cluster [14]. Both these calculations revealed a very strong *s*–*p* hybridization inside the BO_3 group. The top of the filled valence band (ε_v) is formed predominantly by the *s*- and *p* states of oxygen, while the bottom of the empty conduction band (ε_c) is formed primarily by the *s*- and *p* states of boron. The bandgap $E_{g0} = \varepsilon_c - \varepsilon_v$ was estimated at 2.5 eV [13], which is close to the optical absorption edge (2.9 eV [9]). According to the band calculation, the *d* band occurs near the top of the valence band and is partly filled, which corresponds to the metallic state. This drawback of the band theory is related to underestimation of the role of strong electron correlations.

The existing *ab initio* methods do not allow the effects of strong electron correlations to be adequately described. For this reason, we have calculated the structure of *d* bands using the generalized tight binding method [15], in which the addition of electron is related to $d^5 \longrightarrow d^6$ excitation. The energies of d^4, d^5, and d^6 terms are calculated taking into account strong intraatomic Coulomb interactions and are expressed in terms of the Racah parameters $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and the crystal field parameter Δ determined as a difference of the energies of *d* levels (t_{2g} and e_g) in the cubic environment. The energies of the lower terms (corresponding to the high-spin states at normal pressure) are

$$E_0(d^4) \equiv E(^5E, d^4) = 4\varepsilon_d + 6\boldsymbol{A}_4 - 21\boldsymbol{B}_4 - 0.6\Delta_4,$$

$$E_0(d^5) \equiv E(^6A_1, d^5) = 5\varepsilon_d + 10\boldsymbol{A}_5 - 35\boldsymbol{B}_5, \qquad (2)$$

$$E_0(d^6) \equiv E(^5T_2, d^6) = 6\varepsilon_d + 15\boldsymbol{A}_6 - 21\boldsymbol{B}_6 - 0.4\Delta_6,$$

where ε_d is the energy of *d* electrons in a given atom in the one-electron approximation. In the general case, the Racah parameters corresponding to different *n* can vary within approximately 10%. Since our aim here is to understand qualitatively the nature of pressure-induced changes in the spectra and the electron structure, this variation of the Racah parameters can be ignored. The energy of electron addition (analogous to the upper Hubbard band) is

$$\Omega_c = E_0(d^6) - E_0(d^5) = \varepsilon_d + 5\boldsymbol{A} + 14\boldsymbol{B} - 0.4\Delta, \quad (3)$$

and the energy of electron annihilation (or hole creation) is

$$\Omega_v = E_0(d^5) - E_0(d^4) = \varepsilon_d + 4\boldsymbol{A} - 14\boldsymbol{B} + 0.6\Delta. \quad (4)$$

The Mott–Hubbard gap $\Omega_c - \Omega_v \propto U_{\text{eff}} = \boldsymbol{A} + 28\boldsymbol{B} - \Delta$ has the meaning of an effective Coulomb parameter determining the magnitude of strong electron correlations.

In this theory, the parameters are $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, Δ, and $\varepsilon_d - \varepsilon_v$. The values of $\boldsymbol{B} = 0.084$, $\boldsymbol{C} = 0.39$, $\Delta = 1.57$ eV were determined from the optical absorption spectra [6, 7] with the aid of the Tanabe–Sugano diagram [16]. As can be seen from Fig. 3, the exciton transition energies (corresponding to peaks A, B, and C in the experimental spectrum) for these parameters are $\omega_A = 1.39$ eV, $\omega_B = 2.03$ eV, and $\omega_C = 2.80$ eV. The Racah parameter $\boldsymbol{A}$ was determined in [6] from the conditions of coincidence of the energy of excitation with charge transfer $\Omega_c - \varepsilon_v$ with the energy of peak C: $\Omega_c - \varepsilon_v = \omega_C$; the parameter $\varepsilon_d - \varepsilon_v$ was found from the conditions of coincidence of the energy $\Omega_v - \varepsilon_v$ with the energy of the *d* peak in the X-ray spectrum: $\Omega_v - \varepsilon_v = -1.4$ eV [13]. These conditions yield $\boldsymbol{A} = 3.42$ eV and $\varepsilon_d - \varepsilon_v = -14.84$ eV. In what follows, the one-particle energies are measured from the top of the valence band of oxygen states (i.e., $\varepsilon_v = 0$), so that $\varepsilon_d = -14.84$ eV.

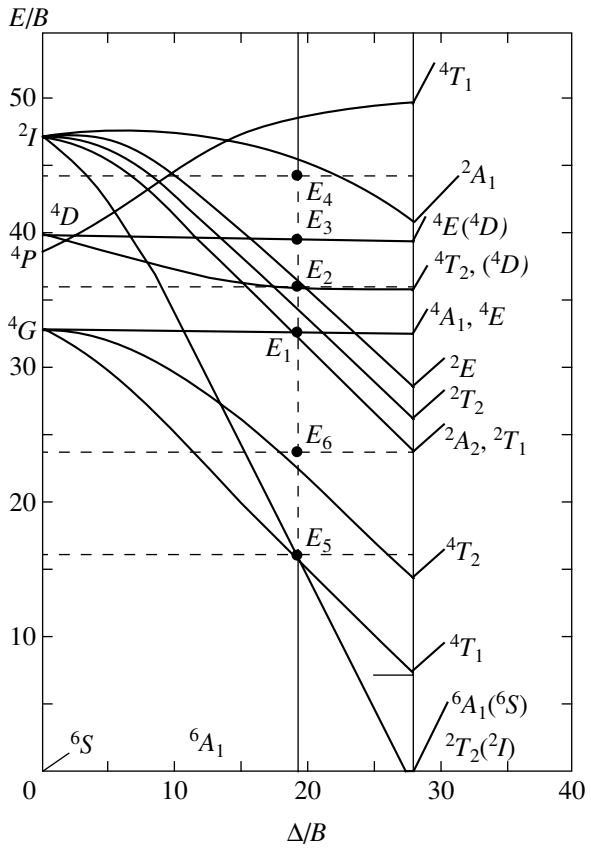

Fig. 3. Tanabe–Sugano diagrams for Fe^{3+} ion. Solid curves show various terms versus the crystal field parameter; dashed lines correspond to the set of parameters $\boldsymbol{B}$, $\boldsymbol{C}$, and Δ for single crystal $FeBO_3$.

Thus, at normal pressure, $U_{\text{eff}} = 4.2$ eV and the electron structure of $FeBO_3$ is characteristic of a charge transfer insulator (see the diagram for $P < P_{cr}$ in Fig. 4).

Other parameters in the theory under consideration is the interatomic hopping parameter *t*, determining the *d*-band halfwidth $W = Zt$ (where $Z = 6$ is the number of nearest neighbors), and the integral of the Fe–Fe exchange interaction $J = 2t^2/U_{\text{eff}}$. The latter quantity, in

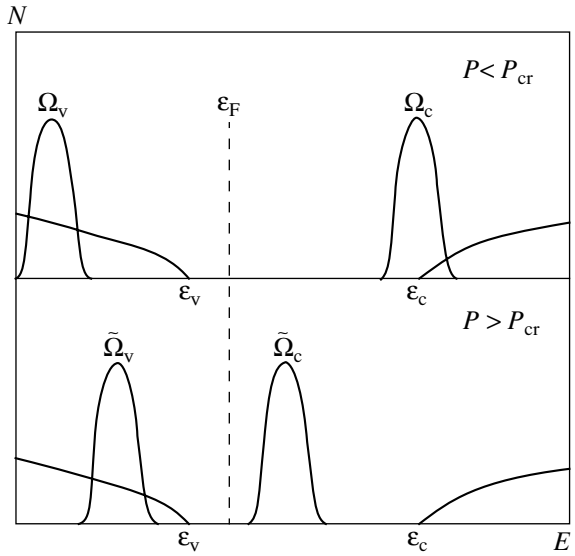

Fig. 4. Density of states in low- and high-pressure phases of single crystal $FeBO_3$ according to the many-electron *p–d* model.

turn, determines the Néel temperature in the mean field approximation: $T_N = ZJS(S+1)/3$. Using these relations and the experimental value of $T_N = 348$ K, we obtain (for $S = 5/2$) $t(P=0) \equiv t_0 = 0.076$ eV. The interatomic *d–d* hopping smears the contributions due to atomlike states (3) and (4) to the total density of states (Fig. 4), so that these contributions change from δ-like to narrow *d* bands.

Now let us consider the pressure-induced changes in the electron structure of $FeBO_3$. The intraatomic Racah parameters $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ can be considered independent of the pressure, whereas the quantities Δ, *t*, and $E_{g0} = \varepsilon_c - \varepsilon_v$ may depend on the interatomic distances. In view of tight binding inside the BO_3 group, we assume that this group is rigid (i.e., insignificantly deformed by increasing pressure). Then, the energy of the top of the valence band and the bandgap E_{g0} between the conduction and valence bands can be considered constant. Assuming also that the parameters *t* and Δ linearly increase with the pressure,

$$\begin{aligned} t(P) &= t_0 + \alpha_t P, \\ \Delta(P) &= \Delta_0 + \alpha_\Delta P, \end{aligned} \tag{5}$$

we infer that U_{eff} linearly decreases as

$$U_{eff}(\Delta) = U_0 - \alpha_\Delta P.$$

The coefficient α_Δ can be determined from the condition of crossover of the high-spin ($^6A_1$) and low-spin ($^2T_2$) terms of the Fe^{3+} ion at a critical value (P_{cr}) of the pressure. According to the Tanabe–Sugano diagram (Fig. 3), the critical value of the crystal field parameter is $\Delta_{cr} \approx 28.5\boldsymbol{B} = 2.4$ eV, from which it follows that $\alpha_\Delta = (\Delta_{cr} - \Delta_0)/P_{cr} = 0.018$ eV/GPa. The coefficient α_t can be determined from the experimentally observed increase in T_N with the pressure [2]. Both the buildup of $t(P)$ and the decrease in $U_{eff}(P)$ lead to the linear increase of the exchange integral $J(P)$, so that $T_N(P)$ can be expressed as

$$T_N(P)/T_N(0) = 1 + (2\alpha_t/t_0 + \alpha_\Delta/U_0)P. \tag{6}$$

According to the experimental data [2], $T_N(0) = 348$ K and $T_N(P_{cr}) \approx 600$ K, this yields $\alpha_t = 0.00046$ eV/GPa. Taking into account the above estimates of $U_0 = 4.2$ eV and $\alpha_\Delta = 0.018$ eV/GPa, we can evaluate the contributions to T_N due to the band broadening (192 K) and due to the decrease in U_{eff} (64 K). Thus, the contribution due to the band broadening is three times that due to the decrease in electron correlations. Nevertheless, the *d*-band halfwidth $W = Zt$ increases rather slightly with pressure: from 0.46 eV at $P = 0$ to 0.58 eV at $P = P_{cr}$.

As a result of the crossover of the ground-state terms of the d^5 and d^4 electron configurations in the high-pressure phase at $P = P_{cr}$, the energies of the lower and upper Hubbard bands exhibit a change. Denoting these energies by $\tilde{\Omega}_v$ and $\tilde{\Omega}_c$,

$$\begin{aligned} \tilde{\Omega}_v &= E(^2T_2, d^5) - E(^3T_1, d^4), \\ \tilde{\Omega}_c &= E(^1A_1, d^6) - E(^2T_2, d^5), \end{aligned} \tag{7}$$

we obtain an expression for the effective Hubbard parameter $\tilde{U}_{eff}$:

$$\tilde{U}_{eff} = \tilde{\Omega}_c - \tilde{\Omega}_v = \boldsymbol{A} + 9\boldsymbol{B} - 7\boldsymbol{C}. \tag{8}$$

Thus, the parameter $\tilde{U}_{eff}$ exhibits a jumplike change at the point of transition, whereby the value at $P > P_{cr}$ ($\tilde{U}_{eff} = 1.45$ eV) is almost one-third of that at $P = 0$. Note that the jump in $\tilde{U}_{eff}$ is only due to the crossover of levels and is not related to the structural transition in $FeBO_3$ [4].

Figure 4 shows the one-electron density of states for both low- and high-pressure phases. As was indicated above, the energies are measured from the top of the valence band of oxygen states (i.e., $\varepsilon_v = 0$). Then, the bottom of the conduction band corresponds to $\varepsilon_c = 2.9$ eV. At a zero applied pressure, we have $\Omega_v = -1.4$ eV and $\Omega_c = 2.8$ eV. Figure 5 shows the pressure dependence of the energies ε_v, ε_c, Ω_v, and Ω_c in the case of pressure-independent Racah parameters and the crystal field parameter Δ linearly increasing with the pressure. Here, we also ignore the possible discontinuity in Δ at the point of the structural phase transition. This jump may produce an additional small shift of the $\tilde{\Omega}_c$ and $\tilde{\Omega}_v$ bands downward, but $\tilde{U}_{eff}$ in the high-pres-

sure phase is independent of P. The smallness of the jump in Δ follows from the experimental fact that the lattice parameter c changes for the most part gradually in the entire interval $0 < P < P_{cr}$ rather than by a jump at P_{cr} [4].

According to Fig. 4, single crystal $FeBO_3$ in the range of $P < P_{cr}$ belongs to the insulators with charge transfer. At $P > P_{cr}$, the insulator character is retained, but the dielectric gap decreases to $E_g = \tilde{\Omega}_c - W_d - \varepsilon_v$. For $P \geq P_{cr}$, this leads to $E_g \approx 0.56$ eV, which allows the material to be considered as a semiconductor. This energy must also determine the edge of optical absorption related to the charge transfer processes. The maximum of absorption is expected at $\tilde{\Omega}_c - \varepsilon_v \approx 0.85$ eV. A sharp decrease in the absorption edge at pressures above P_{cr} is confirmed by the experimental data in Figs. 1 and 2. This drop in the dielectric gap at $P = P_{cr}$ indicates that the electronic transition at this pressure is accompanied by the insulator–semiconductor transition.

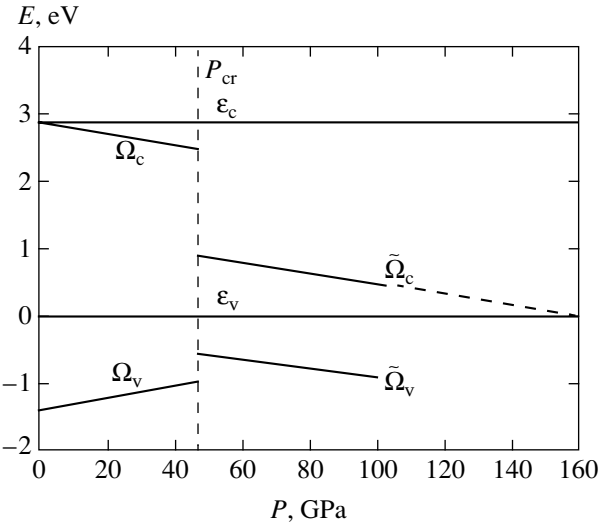

Fig. 5. Shifts of the lower and upper Hubbard bands depending on the pressure in low- and high-pressure phases of single crystal $FeBO_3$.

5. EFFECT OF PRESSURE ON THE OPTICAL ABSORPTION SPECTRUM

Since the pressure dependence of parameters $\Delta(P)$ and $t(P)$ was determined above from independent (nonoptical) data, calculations of the energies of peaks A, B, and C in the optical spectra as functions of the pressure contain no fitting parameters. This provides for the possibility of independent verification of correctness of the proposed model of the electron structure of $FeBO_3$ and the effects of pressure predicted by this model.

According to the Tanabe–Sugano diagram, the energies of $^4T_1$ and $^4T_2$ terms decrease almost linearly with increasing crystal field parameter Δ and, at the critical point, $E(^4T_1, P_{cr}) = 8\boldsymbol{B} \approx 0.67$ eV and $E(^4T_2, P_{cr}) = 15.5\boldsymbol{B} \approx 1.30$ eV. As a result, band A shifts with increasing pressure as $d\omega_A/dP = -0.0156$ eV/GPa. This can be compared to the experimental pressure-induced decrease in the energy of peak A (Fig. 1b): $(d\omega_A/dP)_{exp} = -0.0156$ eV/GPa (see table). For band B, we obtain $(d\omega_B/dP)_{theor} = -0.0158$ eV/GPa.

The energy of $^4A_1$ term for Fe^{3+}, determining the d–d exciton energy ω_C, is independent of the parameter Δ. This agrees well with a weak pressure dependence of the energy of band C observed in experiment (Fig. 1). The energy of excitations with charge transfer, Ω_C, slightly decreases with increasing pressure as $d\Omega_C/dP = -0.4(d\Delta/dP) = -0.0072$ eV/GPa. At $P = P_{cr}$ this yields $\Delta\Omega_C = -0.33$ eV. Probably, it is this small shift that is manifested by a low-energy shoulder observed on band C at high pressures. The most pronounced pressure-induced change in the spectra of a low-pressure phase is the separation of peak C1 from band C. The energy of peak C1 varies most significantly with pressure (Fig. 2). Indeed, according to the Tanabe–Sugano diagram, an increase in pressure in the region of $P = 0$ is accompanied by the crossover of excited terms $^4A_1$ and $^2A_2$ in $FeBO_3$, which is manifested by a new exciton band at $\omega_{C1} = E(^2A_2) - E(^6A_1)$. Since this d–d transition involves a change in the spin ($\Delta S = 2$), the corresponding intensity is small. At $P = P_{cr}$, experiment (Fig. 2) shows $\omega_{C1}(P_{cr}) = 2.0$ eV and, hence, $d\omega_{C1}/dP = -0.0174$ eV/GPa. This is close to the experimental value –0.0199 eV/GPa (see table).

An analogous crossover of terms $^4T_1$ and $^2T_2$ with increasing pressure must lead to an exciton with $\omega_0 = E(^2T_2) - E(^6A_1)$. This exciton must rapidly shift toward infrared region as $d\omega_0/dP = -0.030$ eV/GPa, so that $\omega_0 \approx 0.6$ eV at $P = 20$ GPa. Being doubly forbidden with respect to spin, a ω_0 exciton, as well as ω_{C1}, is expected to possess a low intensity. At $P = P_{cr}$, the energy ω_0 tends to zero. In other words, this exciton plays the role of a "soft" mode for the pressure-induced electronic phase transition.

6. PRESSURE-INDUCED TRANSITION: DISCUSSION OF THE MECHANISM

The observed suppression of magnetic properties at high pressures can be explained a priori using various mechanisms. For example, pressure-induced violation of the local symmetry might lead to the rotation of electron orbitals responsible for the indirect exchange, with the resulting decrease in the indirect exchange interaction. However, as was pointed out in Section 2, in our case the pressure is quasi-hydrostatic and the local symmetry of Fe^{3+} can be considered unchanged. In addition, the observed increase in the Néel temperature with increasing pressure is indicative of enhanced (rather than reduced) exchange interaction.

An alternative approach to explanation of the magnetic collapse in monoxides (FeO, MnO, CoO, and NiO) was suggested by Cohen *et al.* [17] within the framework of *ab initio* calculations performed in the local density and generalized gradient approximations using the generalized Stoner model. It was concluded that the dominating role is played by the *d*-band broadening, which leads to the magnetic collapse and the insulator–metal transition. Indeed, a decrease in the interatomic distances leads to an increase in the *d–d* hopping parameter *t*, so that the Mott–Hubbard insulator with $U \gg W = Zt$ at $P = 0$ may pass to the metallic state with increasing pressure when $W \geq U$. It should be noted that this conclusion is based on the Hubbard model in which *d* electrons exhibit no orbital degeneracy.

The model proposed in [5] is essentially a generalization of the Hubbard model with allowance of the real orbital structure of Fe^{3+}, which leads to the possible coexistence of various spin and orbital multiplets. Since the energies of these multiplets depend on the crystal field parameter Δ, a new mechanism appears for the crossover of optical terms as described by the Tanabe–Sugano diagrams. Of course, the *d*-band broadening still influences the electron structure. Estimations presented in Section 4 showed that the *d* bandwidth in $FeBO_3$ rather insignificantly increases with the pressure, and the main mechanism of magnetic collapse is the crossover from high- to low-spin state. As a result, the effective Hubbard parameter U_{eff} decreases, as can be seen from Fig. 5 and formulas (3) and (4), with increasing Δ and *P*. This implies that an increase in the pressure decreases the role of correlations. The *d*-band broadening also contributes to a decrease in the dielectric gap, but this effect is small as compared to the influence of decreasing U_{eff}.

At pressures above P_{cr}, the ground-state terms of d^4, d^5, and d^6 ions are altered, which leads to a jumplike change in the values of U_{eff} and the dielectric gap, that is, to the transition of $FeBO_3$ crystal to a semiconducting state. As the pressure *P* grows further, we may expect a transition to the metallic state to take place. Since the U_{eff} value at $P > P_{\text{cr}}$ no longer depends on the pressure, while the *d* bandwidth linearly increases with *P*, an insulator–metal transition of the Mott–Hubbard type becomes possible. Once the baric derivative of α_t is known, we can readily evaluate the critical Mott–Hubbard pressure P_{MH} at which the complete metallization takes place and $W(P_{\text{MH}}) = \tilde{U}_{\text{eff}} = 1.45$ eV, which yields $P_{\text{MH}} = 360$ GPa. On the other hand, it is interesting to note that extrapolation of the $\tilde{\Omega}_c(P)$ plot in Fig. 5 to the intersection with the top (ε_v) of the valence band yields $P_{\text{met}} = 200$ GPa. This value virtually coincides with the experimental estimate of the pressure $P_{\text{met}} = 210$ GPa for the complete metallization, obtained by extrapolating the thermoactivation gap to zero [3]. However, the transition to the metallic state in this model should be expected at lower pressures, since the E_g value tends to zero not only due to a decrease in $\tilde{\Omega}_c$, but also due to an increase in the *d* bandwidth W_d (additionally decreasing the gap). In such cases, a transition to the metallic state is caused merely by the crossover of bands due to an increase in the pressure, since the Fermi level must occur below the top of the valence band and the Fermi surface will open. According to Lifshits *et al.* [18], this process is classified as the 2.5-order transition. Taking into account the baric derivatives of α_t and α_Δ determined from the optical and magnetic data for the low-pressure phase, we can estimate the metallization pressure as $P_{\text{met}} \approx 73$ GPa. However, experiments [3] show that $FeBO_3$ at this pressure still occurs in a semiconducting state. Apparently, extrapolation of the baric derivatives of α_t and α_Δ determined for the low-pressure phase to the region of high pressures does not provide for sufficiently accurate evaluation.

7. CONCLUSIONS

The band structure and optical spectra of the Mott–Hubbard insulators (including $FeBO_3$) must depend on the pressure mostly for two reasons: (i) a growth in the pressure may give rise to the *d* bandwidth due to an increase in the integral of interatomic electron hopping and (ii) pressure variations alter the crystal field. The results of our calculations taking into account peculiarities of the crystal structure of $FeBO_3$ lead to a conclusion that a determining role is played by the pressure-induced change in the crystal field parameter Δ. The main peculiarity in the crystal structure of $FeBO_3$ is very strong *s–p* hybridization inside the BO_3 group, which leads to a very weak *p–d* hybridization between oxygen and iron ions and to a small width of the *d* band. The *d* band exhibits additional narrowing in the antiferromagnetic phase due to the spin–polaron interaction.

The increase in Δ not only leads to the crossover of high- and low-spin states of Fe^{3+} ion, explaining the collapse of the magnetic moment, but also causes an analogous crossover of the Fe^{2+} and Fe^{4+} configurations. As a result, the effective Hubbard parameter

$$U_{\text{eff}} = E_0(d^4) + E_0(d^6) - 2E_0(d^5),$$

which is a measure of the Coulomb correlations, is determined in the low-pressure phase by the high-spin terms of the ground states of d^4, d^5, and d^6 configurations, and in the high-pressure phase, by the corresponding low-spin terms, which accounts for a jumplike decrease in U_{eff} at the point of transition. Although U_{eff} decreases by a factor of almost three, the metallization does not take place because of a small *d* bandwidth. As the pressure grows further, a transition of $FeBO_3$ to the metallic state unavoidably takes place. It is interesting to note that extrapolation of the $\tilde{\Omega}_c$ level to inter-

section with the top of the valence band allows the pressure required for a transition to the metallic state to be estimated as $P_{met} = 200$ GPa. This estimate nearly coincides with the experimental value evaluated from the condition of zero thermoactivation gap [3]. However, such extrapolation fully ignores the contribution of the *d*-band broadening to a decrease in the dielectric gap. On the other hand, extrapolation of the pressure-induced band broadening determined for the low-pressure phase to the region of high pressures gives an obviously understated value of $P_{met} = 73$ GPa. Apparently, the *d* bandwidth in the high-pressure phase grows with the pressure slower than in the low-pressure phase. It is also possible that the top of the valence band slightly decreases with increasing pressure, which leads to an increase in P_{met}.

Within the framework of our analysis of the pressure dependence of the positions of optical absorption bands (in the low-pressure phase), let us compare the experimental and theoretical data presented in the third and fifth columns of the table, respectively. As can be seen, there is qualitative agreement for all bands in the absorption spectrum of $FeBO_3$. The linear variation of the energies of bands A and B with the pressure confirms the hypothesis of a linear pressure dependence of the crystal field (formula (1)). At the same time, we would like to point out that there is no quantitative agreement between theory and experiment for bands A and B, and the reasons for this discrepancy are unclear. The most unusual feature in the behavior of experimental optical spectra is the separation of peak C1 (most significantly changing with the pressure) from band C. For this peak, found to be related to the ${}^6A_1 \longrightarrow {}^2A_2$ exciton, we observed the best agreement between theory and experiment. On the other hand, it is unclear why the expected exciton ${}^6A_1 \longrightarrow {}^2T_2$ with an energy below that of band A is not observed. This very exciton must play the role of a soft mode for the pressure-induced electronic phase transition, since the transition is due to the crossover of ${}^6A_1$ and ${}^2T_2$ terms of Fe^{3+} ion.

In conclusion, it should be noted that the proposed many-electron model of the electron structure of $FeBO_3$ taking into account strong electron correlations not only qualitatively describes the very fact of the pressure-induced electronic and magnetic transitions, but also explains fine experimental details such as variations in the optical spectra that depend on the pressure.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project nos. 02-02-17364a, 03-02-16286a, and 04-02-16945a) and the program "Strongly Correlated Electrons" of the Department of Physical Sciences of the Russian Academy of Sciences.

REFERENCES

1. I. A. Troyan, A. G. Gavriliuk, V. A. Sarkisyan, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **74**, 26 (2001) [JETP Lett. **74**, 24 (2001)].
2. V. A. Sarkisyan, I. A. Troyan, I. S. Lyubutin, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 788 (2002) [JETP Lett. **76**, 664 (2002)].
3. I. A. Troyan, M. I. Eremets, A. G. Gavriliuk, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **78**, 16 (2003) [JETP Lett. **78**, 13 (2003)].
4. A. G. Gavriliuk, I. A. Trojan, R. Boehler, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **75**, 25 (2002) [JETP Lett. **75**, 23 (2002)].
5. S. G. Ovchinnikov, Pis'ma Zh. Éksp. Teor. Fiz. **77**, 808 (2003) [JETP Lett. **77**, 676 (2003)].
6. S. G. Ovchinnikov and V. N. Zabluda, Zh. Éksp. Teor. Fiz. **125**, 198 (2004) [JETP **98**, 135 (2004)].
7. A. V. Malakhovskii and I. S. Edelman, Phys. Status Solidi B **74**, K145 (1976).
8. A. J. Kurzig, R. Wolte, R. C. Graw, and J. W. Nielsen, Appl. Phys. Lett. **14**, 350 (1969).
9. I. S. Édel'man, A. V. Malakhovskiĭ, T. I. Vasil'eva, and V. N. Seleznev, Fiz. Tverd. Tela (Leningrad) **14**, 2810 (1972) [Sov. Phys. Solid State **14**, 2442 (1972)].
10. B. Andlauer, D. F. Schirmer, and J. Schneider, Solid State Commun. **13**, 1655 (1973).
11. A. V. Malakhovskii and I. S. Edelman, Solid State Commun. **28**, 475 (1978).
12. V. N. Zabluda and A. V. Malakhovskiĭ, Zh. Éksp. Teor. Fiz. **121**, 133 (1985) [Sov. Phys. JETP **27**, 77 (1985)].
13. A. V. Postnikov, St. Bartkovski, M. Newmann, *et al.*, Phys. Rev. B **50**, 14849 (1994).
14. N. B. Ivanova, V. V. Rudenko, A. D. Balaev, *et al.*, Zh. Éksp. Teor. Fiz. **121**, 354 (2002) [JETP **94**, 299 (2002)].
15. V. V. Val'kov and S. G. Ovchinnikov, *Quasi-Particles in Strongly Correlated Systems* (Sib. Otd. Ross. Akad. Nauk, Novosibirsk, 2001) [in Russian].
16. Y. Tanabe and S. Sugano, J. Phys. Soc. Jpn. **9**, 766 (1954).
17. R. E. Cohen, I. I. Mazin, and D. G. Isaak, Science **275**, 654 (1997).
18. I. M. Lifshits, M. Ya. Azbel', and M. I. Kaganov, *Electron Theory of Metals* (Nauka, Moscow, 1971; Consultants Bureau, New York, 1973).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 3, 2004, pp. 574–584.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 3, 2004, pp. 659–671.

SOLIDS
Electronic Properties

DX^- Center Formation in Planar-Doped GaAs:Si in Strong Electric Fields

M. Asche[a,*] and O. G. Sarbey[b,]**
[a]*Greifswaldestrasse 88, 10409 Berlin, Germany*
**e-mail: marion.asche@freenet.de*
[b]*Institute of Physics, National Academy of Sciences of Ukraine, 03028 Kiev, Ukraine*
***e-mail: sarbey@profit.net.ua*
Received December 18, 2002; in final form, January 20, 2004

Abstract—The kinetics of current decay and partial restoration in planar doped GaAs:Si due to the formation of DX^- centers in strong electric fields has been experimentally studied. The existence of thresholds with respect to the field strength and donor concentration is explained. A model of the DX^- center formation is proposed, which is based on the notions about variation of the depth and width of a potential well created by planar doping, caused by the redistribution of hot electrons between quantum confinement subbands. As a result, the energy level of DX^- centers, which is situated above the potential well depth in the absence of strong field, decreases and falls within the potential well. This makes possible the DX^- center formation, provided that hot electrons, occupying the resonance electron levels in the conduction band, simultaneously excite local vibrational modes. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Most experimental and theoretical investigations of the DX^- centers performed up to now have been related to homogeneously doped materials (three-dimensional layers or massive samples), but some experimental (see, e.g., [1–5]) and recent theoretical studies [6–8] were devoted to two-dimensional structures in gallium arsenide (GaAs). From the experimental standpoint, planar doped structures, where very high dopant concentrations can be obtained and very strong electric fields can be simultaneously created without heating the lattice and modifying the sample properties, are very convenient objects for investigating the effect of strong electric fields on the kinetics of DX^- center formation.

Using short-pulse probing techniques, we studied the current as a function of the electric field strength [4, 5] and peculiarities of the phonon time-of-flight spectra [9] in δ-doped GaAs:Si structures. The observed current decay kinetics at a fixed field and features of the time-of-flight spectra were interpreted as manifestations of the appearance of metastable centers formed in the heating electric field.

Previously [10], we attempted to determine some characteristics of the DX^- center formation and ionization in an electric field heating the charge carriers. However, to our knowledge, no clear physical model explaining the effect of such fields on the observed transition from band states to a resonance level has been proposed in the literature. Recently [11], we showed that the heating of electrons in a δ-doped GaAs:Si structure is accompanied by a change in the depth and shape of the potential well. We believe that this effect must play a significant role in the formation of DX^- centers in strong electric fields. Indeed, DX^- center formation requires the electron density near Si atoms to be sufficiently high at energies corresponding to the energy level of these centers. For this purpose, the energy levels of DX^- centers must lie below the energy corresponding to the edge of the well. However, analysis of published data shows that, for average levels of doping and low temperatures at which the formation of DX^- centers in electric fields heating the electrons is observed, this condition is not satisfied.

This study was aimed at obtaining new experimental data through the investigation of current transients in planar-doped GaAs:Si with various concentrations of silicon at liquid nitrogen and liquid helium temperatures. We will also propose and discuss a physical model of the formation of DX^- centers under the action of a strong electric field.

2. EXPERIMENTAL METHODS AND RESULTS

2.1. Sample Preparation and Experimental Procedures

The experiments were performed on GaAs:Si structures grown by molecular beam epitaxy (MBE). In order to determine how can the structural features influence DX^- center formation, we used samples grown using different MBE procedures.

The structure of type 1 was obtained on a GaAs(001) single crystal surface at a constant substrate temperature of 550°C. The δ-layer was grown by continuously supplying silicon atoms at an equivalent partial pressure of arsenic $p = 7.1 \times 10^{-6}$ mbar [4]. The concentration of silicon atoms in this structure was $N_{Si} = 7 \times 10^{12}$ cm^{-2}. The δ-layer featured a 2×3 reconstruction.

The structure of type 2 was grown on a vicinal GaAs substrate with a 2° misorientation from (001) toward (111) gallium plane. In this case, a silicon layer with $N_{Si} = 1.4 \times 10^{14}$ cm^{-2} was deposited in a pulsed regime [12] at 610°C and an equivalent partial pressure of arsenic $p = 6 \times 10^{-6}$ mbar, after which the temperature was reduced to 580°C and a cap layer of GaAs was deposited. For the given concentration of silicon and the substrate orientation, the samples of type 2 are characterized by a 3×2 silicon structure grown on a (2×4) α-reconstructed GaAs surface [12–14].

In the typical mesostructures obtained by photolithography on the samples of types 1 and 2 (see the inset to Fig. 1), the distance between current-carrying contacts was $W = 8.3$ or 19.1 μm and the contact widths was 120 or 180 μm, respectively. The contacts prepared from Au:Ge alloy were burned in at $T = 430$°C and the surface layer was partly scoured by etching prior to metallization [4].

The current–voltage characteristics (*I–U* curves) and the current kinetics at various applied fields were measured in the dark at liquid nitrogen or liquid helium temperature. The voltage was applied in the form of single rectangular pulses with a length of 240 ns (for the *I–U* measurements) or shorter (for the kinetic measurements), or in the form of repeated pulses with the repetition rate gradually increased up to 200 Hz.

The amplitudes of current and voltage pulses were measured with an oscillograph with a high-voltage high-frequency voltage divider. Since resistance of many samples was in the region of 20–40 Ω, the load was matched to the oscillator output by a 50 Ω resistor.

In order to provide for reproducible initial conditions during the investigation of transient kinetics, the samples were heated to room temperature under illumination after each measurement and then slowly cooled in the dark prior to the next run.

2.2. Experimental Results

The electron density n and mobility μ were determined from the Hall effect and conductivity measurements at $T = 77$ K in a magnetic field of about 1 kG. For the structures of types 1 and 2, these values were $n = 3.5 \times 10^{12}$ cm^{-2}, $\mu = 3300$ cm^2/(V s) and $n = 1.5 \times 10^{13}$ cm^{-2}, $\mu = 1240$ cm^2/(V s), respectively. Using the structures of type 1, it was possible to measure the Shubnikov–de Haas effect [4] and determine the electron densities in the first three quantum confinement subbands: 3.3×10^{12}, 1.3×10^{12}, and 0.4×10^{12} cm^{-2}.

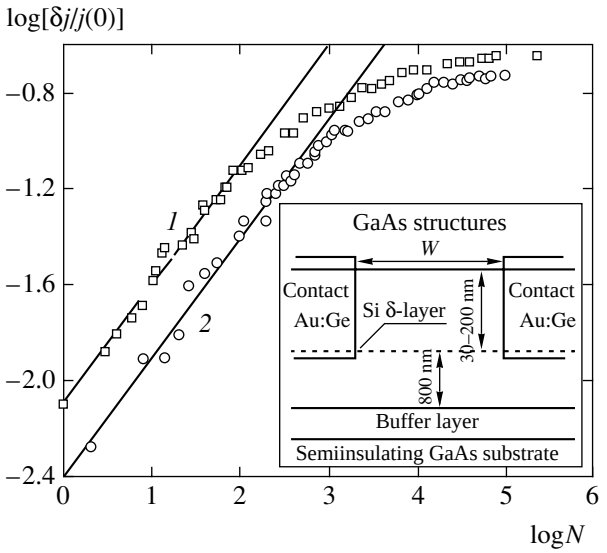

Fig. 1. Double logarithmic plots of the normalized current change versus number of field pulses (proportional to the duration of field action) for planar-doped GaAs:Si structures with $N_{Si} = 7 \times 10^{12}$ (*1*) and 1.4×10^{14} cm^{-2} (*2*) and the initial Hall electron densities $n = 3.5 \times 10^{12}$ (*1*) and 1.5×10^{13} cm^{-2} (*2*). The kinetic curves were measured at $T = 77$ K and a field strength of (*1*) 17.7 and (*2*) 18.8 kV/cm. Symbols represent the experimental data; lines show linear approximation of the initial stage. The inset shows a schematic diagram of the sample.

In high electric fields, the current decreased with the time t of voltage application (i.e., with the number of pulses) in the structures of both types. However, one sample of type 1 with a significantly lower silicon concentration ($N_{Si} = 4 \times 10^{12}$ cm^{-2}) exhibited no variation of the current with time [15]. This fact indicates that there exists a minimum threshold electron density above which the formation of DX⁻ centers is possible.

Figure 1 shows typical curves of the current change $\delta j(t) = j(0) - j(t)$ normalized to the initial value $j(0)$ at $t = 0$ as a function of the number N of voltage pulses, plotted in the double logarithmic scale. In the region of small N, the plots are linear and have a characteristic slope of 0.5, which indicates that the relative change of the current in this region is proportional to the square root of time.[1] This scale will be used for plotting the kinetic curves below.

Before proceeding with the presentation of experimental data, it is necessary to mention one peculiarity of our experimental procedure. It is obvious that, at a fixed output voltage of the voltage generator, a decrease in the current leads to a decreased voltage drop on the matching resistor and, hence, to an increase in the field

[1] It should be noted that this behavior was observed in most cases for the structures of both types. However, a large scatter of experimental points in some cases did not allow the slope of the logarithmic plots to be unambiguously determined, the possible value ranging between 0.5 and 1.0.

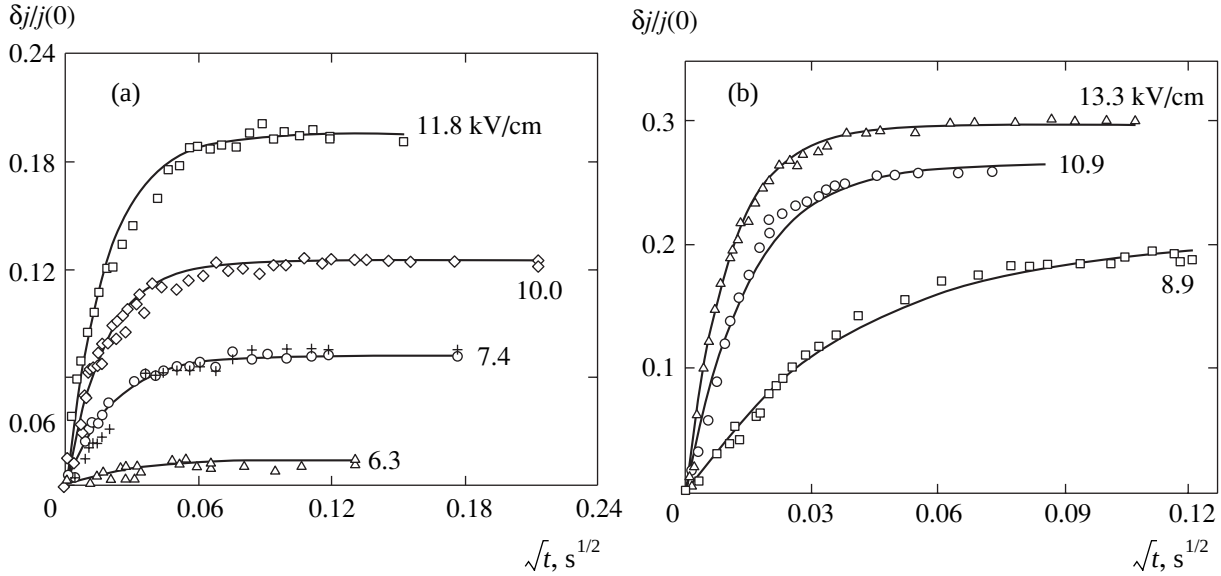

Fig. 2. Plots of the normalized current change versus square root of the duration of field action for a planar-doped GaAs:Si structure of type 2 with $N_{Si} = 1.4 \times 10^{14}$ cm^{-2} and the initial Hall electron density $n = 1.5 \times 10^{13}$ cm^{-2}. The kinetic measurements were performed at $T = 77$ (a) and 4.2 K (b) and various field strength (indicated at the curves). Symbols represent the experimental data; curves show the results of approximation using Eq. (1) for the same field strength as the parameter. Crosses show the results of measurements at a field strength of 7.4 kV/cm plotted without current correction.

strength in the sample. Therefore, the measured current change is related both to the expected decrease in the conductivity and to an increase in the field strength. In order to exclude the second factor, we introduced a correction to the current measured at each given time *t*. The correction was determined using the current–voltage characteristic and the value of voltage at this time.

The results of such correction are illustrated in Fig. 2, presenting the kinetic curves $\delta j(t)/j(0)$ (t is the total time of field action) measured in various fields at 77 and 4.2 K for a sample doped to 1.4×10^{14} cm^{-2}. In agreement with the data in Fig. 1, an increase in the duration of field action leads to deviation of the current change from proportionality to the square root of time and reveals a clear tendency to saturation. It should be noted that the current change in this sample observed in a field of 6.3 kV/cm were as small as about 1%.

A significant decrease in the maximum change of current (for a large number of pulses $N \sim 10^6$) at certain field strengths was observed in all samples. This result suggests that the phenomenon under consideration is characterized by a minimum threshold with respect to the field strength.

To the best of our knowledge, there is no theory to which the obtained could be compared. In order to provide for a quantitative description of data, we have approximated the experimental points using an expression taking into account both characteristic features of the observed behavior: (i) proportionality to the square root of time in the initial region and (ii) saturation for large times at a level dependent on the field strength. For fixed field strength, a simple expression meeting both conditions is the two-parametric formula [16]

$$\frac{\delta j(t)}{j(0)} = \frac{\delta j_s(t)}{j(0)}\left[1 - \exp\left(-\sqrt{\frac{t}{\tau}}\right)\right], \qquad (1)$$

where δj_s is the value of δj in the region of saturation and

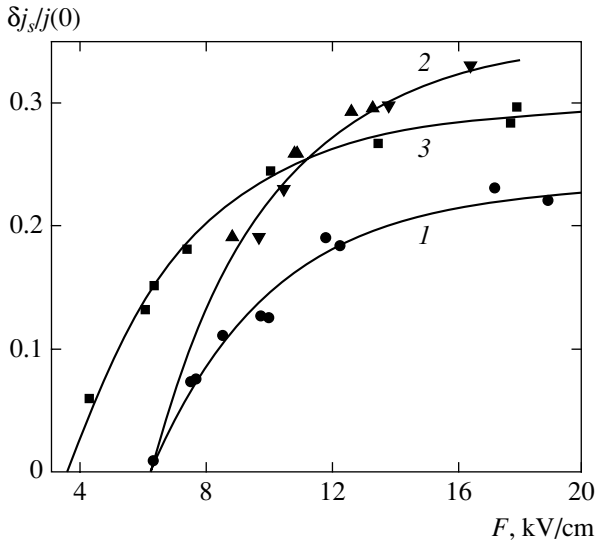

Fig. 3. Plots of the normalized current change at saturation versus field strength F for GaAs:Si structures of (*1*, *2*) type 2 with $N_{Si} = 1.4 \times 10^{14}$ cm^{-2} measured at (*1*) 77 and (*2*, data for two samples) 4.2 K and for a sample of (*3*) type 1 with $N_{Si} = 7 \times 10^{12}$ cm^{-2} at 77 K. Symbols represent the experimental data; curves show the results of approximation using Eq. (2) with the parameters indicated in the table.

τ is the characteristic time of current decay. Depending on the particular sample and the applied field strength, the experimental points could be more or less satisfactorily described in terms of this expression. The corresponding curves are depicted in Fig. 2 by solid curves.

Figure 3 presents the curves of $\delta j_s/j(0)$ versus field strength measured at 77 and 4.2 K for the same sample of type 2 ($N_{Si} = 1.4 \times 10^{14}$ cm^{-2}) as in Fig. 2 and the curve measured at 77 K for a sample of type 1 ($N_{Si} = 7 \times 10^{12}$ cm^{-2}). In order to demonstrate the reproducibility of results, we also present the data for another samples of type 2 measured at 4.2 K. The curves in Fig. 3 clearly reveal the presence of thresholds and show saturation of the $\delta j_s/j(0)$ value at large field strengths. It should be emphasized that the level of saturation, although markedly below unity, is nevertheless relatively large.

Proceeding from considerations analogous to those used in obtaining formula (1), we have approximated the experimental data in Fig. 3 using a three-parametric formula

$$\frac{\delta j_s(t)}{j(0)} = \frac{\delta j_{ss}(t)}{j(0)}\left[1 - \exp\left(-\sqrt{\frac{F - F_0}{F_j}}\right)\right], \qquad (2)$$

where F_0 is the threshold field strength, δj_{ss} is the value of δj_s for $F \longrightarrow \infty$, and F_j characterizes the rate of saturation current increase with the field strength. The curves calculated using formula (2) are depicted by solid lines in Fig. 3; the parameters of approximating curves are listed in the table. A comparison of these parameters shows that the maximum current change decreases with increasing silicon content and temperature. The threshold field increases with the doping level and rather weakly depends on the temperature (at least, in the temperature range studied). It is noteworthy that all data presented in Fig. 3 can be described by formula (2) with the same F_j (to within the given approximation accuracy).

A considerable scatter of τ values for various field strength usually hinders unambiguous interpretation of this dependence. Nevertheless, it was possible to estimate the characteristic time, which ranges from several milliseconds to several tenths of a millisecond and sharply drops when the field strength increases from the threshold to 8–10 kV/cm. It is also possible to ascertain that the current varies with time more rapidly at lower temperatures and higher doping levels.

Some of the samples exhibited an interesting peculiarity, which was manifested at the maximum field strengths used in our experiments. In these cases, the current changed in a stepwise manner and the experimental data could not be described by simple formula (1). However, a description could be provided by using two or three such expressions with different values of $\delta j_s/j(0)$ and τ for various time intervals. This situation is illustrated in Fig. 4. We believe that this behavior is related to the well-known phenomenon of so-called high-field domain formation in GaAs exposed to high electric fields.

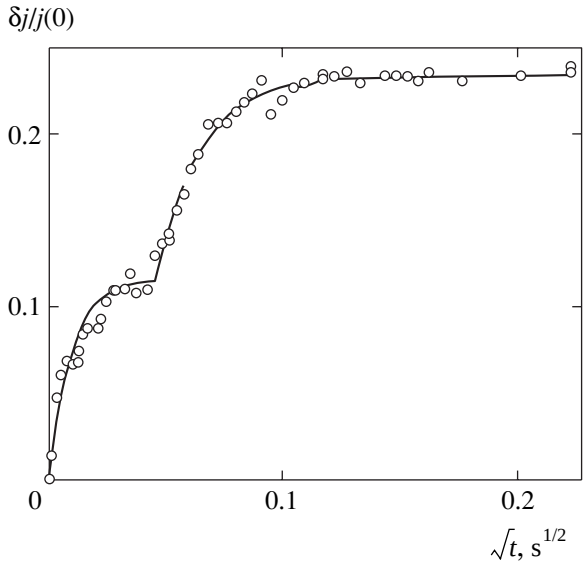

Fig. 4. Plot of the normalized current change versus square root of the duration of field action for a planar-doped GaAs:Si structure of type 2 with $N_{Si} = 1.4 \times 10^{14}$ cm^{-2} in the presence of a high-field domain. The kinetic measurements were performed at $T = 77$ K and a field strength of $F = 12.7$ kV/cm.

As indicated previously [4], the conductivity of a GaAs:Si structure remains unchanged for several days after switching off the electric field, provided that the samples are kept at low temperature in the dark. If a sample is exposed to a strong electric field until the current change exhibits saturation and then the field strength is decreased, then the current increases to reach a value corresponding to the new field. Figure 5 shows how a change in the current saturated as a result of exposure in a strong electric field depends on a weaker "destructing" field. This partial restoration effect suggests that the saturation current and the field strength are uniquely interrelated and that the former value depends only on the experimental conditions and the field strength rather than on the sample prehistory.

3. DISCUSSION

3.1. Dopant Layer Structure

For the correct interpretation of experimental data, it is necessary to answer the question as to whether our structures can be considered as Si–δ-doped. In order to answer this question, let us consider the distribution of

Table

n, 10^{12}, cm^{-2}	T, K	$\delta j_s/j(0)$	F_0, kV/cm	F_j, kV/cm
3.5	77	0.295	3.60	3.9
15	77	0.230	6.15	3.9
15	4.2	0.350	6.20	3.9

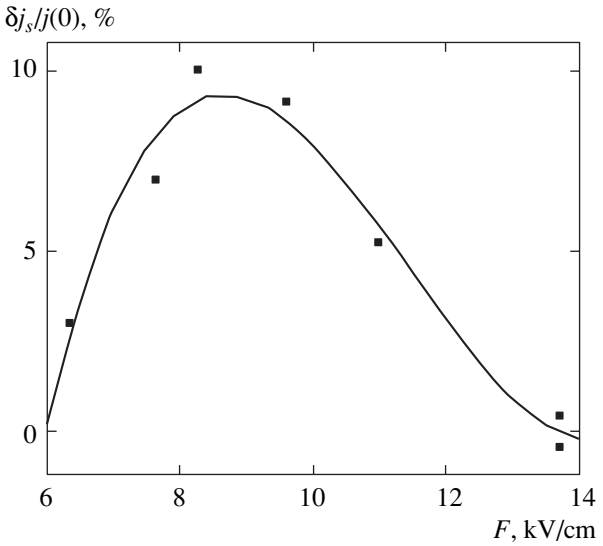

Fig. 5. Plot of the normalized current change at saturation versus field strength F after preliminary decrease of the current at F_s = 13.8 kV/cm and T = 77 K.

Si atoms in the vicinity of the plane of doping. This distribution strongly depends on the temperature and other growth conditions and significantly varies in various investigations. However, an analysis of the published data [1, 2, 17–20] allowed us to conclude that "smearing" of the doped layer in the structures of type 1 should be small (FWHM within 4 nm) as compared to the width of the electron distribution (above 10 nm).

The situation with the GaAS:Si structures of type 2 is not as clear. In this case, the crystallographic structure of the growth surface and the growth temperature allow Si atoms to migrate toward the edges of terraces and exhibit predominant ordering in the form of dimmers (see [12–14]). Apparently, this order only slightly degrades in the course of subsequent growth [21]. For this reason, we may suggest that, despite relatively high growth temperature, the narrow distribution of silicon is retained, provided that the temperature is decreased immediately after formation of the doped layer. At the same time, a small fraction of Si atoms, which do not form dimers, can diffuse at the growth temperature through a depth of up to 20 nm.

Another important fact is that the concentration of silicon atoms and the electron density determined from the Hall effect measurements usually [1, 19, 22] differ by a factor of two in the structures of type 1 and by a factor of ten or above in highly doped structures of type 2. As is known, the Hall measurements fail to give correct values of the total carrier density because of different electron mobilities in various subbands of the Si–δ-doped structure. The results of measurements of the Shubnikov–de Haas effect in the samples of type 1 mentioned above showed that the uncertainty falls within 30%. However, even this correction does not eliminate the discrepancy, which is especially pronounced in the structures with greater concentrations of silicon atoms.

For considering this problem, it is necessary to know the distribution of silicon atoms in the matrix lattice. The results of various investigations of the homogeneously doped GaAs indicate that Si atoms may be incorporated into the lattice not only at cation sites (to form shallow-donor Si_{Ga} levels), but also at anion sites (with the formation of Si_{As}^{-} acceptors) and as components of the Si_{Ga}^{+}–Si_{As}^{-} and Si–V_{Ga} complexes. In the latter case, the V_{Ga} defect can exist in the form of V_{Ga}^{-}, V_{Ga}^{-2}, V_{Ga}^{-3}, or Si–X complex assigned to a flat V_{Ga}–Si_{As}–As_{Ga} complex but probably representing a disturbed variant of the vacancy complex [23, 24]. The Raman spectra of local vibrational modes show that such defects are also present in our samples [4, 13].

Assuming that the scattering cross sections of these defects are on the same order as in [25], it is possible to eliminate the discrepancy between the Hall effect data and the dopant concentration only for the samples of type 1 (doped to a lower level).

In the case of GaAs homogeneously doped to a high level (>10^{19} cm^{-3}), the formation of DX^{-} centers may become a significant factor, leading to a decrease in the density of free carriers relative to the concentration of donors [26]. It was suggested that the discrepancy observed in planar-doped structures can be also related to the presence of DX^{-} centers [1,22, 23, 27]; Arscott *et al.* [28] claimed to have found direct evidence for this effect in a sample with $N_{Si} = 10^{13}$ cm^{-2}.

This situation requires certain comments. In hydrostatically compressed homogeneously doped GaAs samples with $N_{Si} = 6.3 \times 10^{17}$ cm^{-3}, the absorption spectrum of local vibrational modes exhibited a peak at λ = 395 cm^{-1} [29]. This signal was assigned to DX^{-} centers and theoretically considered in [30, 31]. However, such peaks (extrapolated approximately to λ = 376 cm^{-1} at atmospheric pressure) were not observed in the IR absorption spectra of local vibrational modes for Si-δ-doped structures [22] or in the Raman spectra of our structures [4, 13] to within the sensitivity of the methods employed. Moreover, the main feature of DX^{-} centers manifested by the existence of frozen photoconductivity at low temperatures was not unambiguously established in [1, 2, 32]. For these reasons, we believe that the concentration of DX^{-} centers in our structures in the initial state is negligibly small (or such centers are absent).

In order to explain a large difference between the free carrier density and the dopant concentration in the structures of type 2, we assume that electrons are bound on silicon atoms ordered in pairs occupying the neighboring gallium sites. It is still unclear whether these pairs are like dimers present on a growth surface with partial charge transfer [33–36] or represent complexes

of covalently bound silicon atoms of the Si_{Ga}–Si_{Ga} type suggested and theoretically studied in [37]. It should be noted that an electrically neutral doped surface was actually observed in [38]. If this is true, then only separate Si atoms situated in the vicinity of the doping plane can supply electrons to the conduction band.

3.2. Calculation of the Energy Spectrum

In order to interpret the experimental data, we calculated the energy spectrum of the Si-δ–doped layer at various temperatures using the method described in [39]. The calculation was performed for a model comprising the ideal chain of independent quantum wells. For this model to be applicable, it was necessary that, on the one hand, the distance between quantum wells would be sufficiently large and the overlap between electron work functions even for high levels in the quantum well could be ignored. On the other hand, this distance must be not too large, so that one doped plane would account for a small number of states with the energies greater than that of the well edge. Under this condition, the occupancy of these states can be ignored. For the electron densities and temperatures used in our experiments, both conditions can be satisfied if the distance between δ-layers falls within 150–200 nm. We ignore nonparabolic shape of the conduction band. The influence of this factor on the positions of energy levels and the occupancy of states does not exceed 10% for the maximum electron density ($n = 1.5 \times 10^{13}$ cm^{-2}) used in the calculations [40]. The positive charge of silicon ions is assumed to be homogeneously distributed in the plane (jelly model). This approximation is valid for silicon ion concentration $N_{Si^+} > 10^{12}$ cm^{-2}. In this case, the average distance between ions is smaller than the characteristic spatial size of the electron wave function even in the ground state.

The system of equations to be solved comprises one-dimensional Schrödinger and Poisson equations supplemented by the condition of quasineutrality [11]. In terms of the dimensionless variables $L_0 = \hbar^2\varepsilon/me^2$, $E_0 = me^4/\varepsilon^2\hbar^2$, $N_0 = 1/L_0^2$, where ε is the permittivity, these equations are as follows:

$$\frac{d^2\psi_n(x)}{dx^2} + 2[E_{n0} - V(x) - V_{ex}]\psi_n(x) = 0, \quad (3)$$

$$\frac{d^2V(x)}{dx^2} = 4\pi\left\{N_s f(s) - \frac{T}{\pi} \times \sum_n \ln\left[1 + \exp\left(-\frac{E_{n0} - \zeta}{T}\right)\right]|\psi_n|^2\right\}, \quad (4)$$

$$N_s - \frac{T}{\pi}\sum_n \ln\left[1 + \exp\left(-\frac{E_{n0} - \zeta}{T}\right)\right] = 0. \quad (5)$$

Here, $V(x)$ is the self-consistent energy of the electron cloud in the field of silicon ions with neglect of the exchange-correlation contribution, V_{ex} is the exchange-correlation energy, ζ is the Fermi level, E_{n0} is the energy of the nth subband bottom, T is the temperature (in dimensionless energy units), $f(x)$ is the distribution function of positive ions (assumed to be δ-shaped or Gaussian), and N_s is the ion density related to the plane. We used the exchange-correlation energy in the local density approximation [41]

$$V_{ex} = -\frac{1}{\alpha\pi r_s} + \gamma\frac{1.16\beta_1\sqrt{r_s} + 1.33\beta_2 r_s}{(1 + \beta_1\sqrt{r_s} + \beta_2 r_s)^2}, \quad r_s \geq 1, \quad (6)$$

$$V_{ex} = -\frac{1}{\alpha\pi r_s} + A\ln r_s + \left(B - \frac{A}{3}\right) + 2Cr_s\ln\left(\frac{r_s}{3}\right) + \frac{1}{3}(2D - C)r_s, \quad r_s < 1, \quad (7)$$

where

$$r_s = \left(\frac{0.75}{n}\right)^{1/3},$$

$$n = \frac{T}{\pi}\sum_n \ln\left[1 - \exp\left(-\frac{E_{n0} - \zeta}{T}\right)\right]|\psi_n|^2,$$

and β_i, A, B, C, and D are constants given in [41]. In these expressions, the exchange-correlation energy does not explicitly depend on the temperature. As demonstrated in [42], this dependence is insignificant in the concentration range of interest and, hence, can be ignored.

The corresponding boundary conditions are as follows:

$$V(\infty) = 0, \quad \psi_n(\infty) = 0,$$

$$\frac{dV(0)}{dx} = 2\pi N_s \text{ (for the δ-shaped distribution function)},$$

$$\psi_n(0) = 0 \text{ (for even } n),$$

$$d\psi_n(0)/dx = 0 \text{ (for odd } n).$$

The system of equations (3)–(5) with the boundary conditions was numerically solved by the iteration method foe various temperatures and dopant concentrations. In the zero approximation, the potential was calculated as a linear combination of two parts. The first part was determined by the electron charge and analytically calculated using the Poisson equation in the Thomas–Fermi approximation and the second part was due to positive ions. The exchange-correlation energy was included into all subsequent approximations.

The calculations were performed for various distances within 150 nm from the plane of doping. The interval between neighboring points was 0.1 nm for the

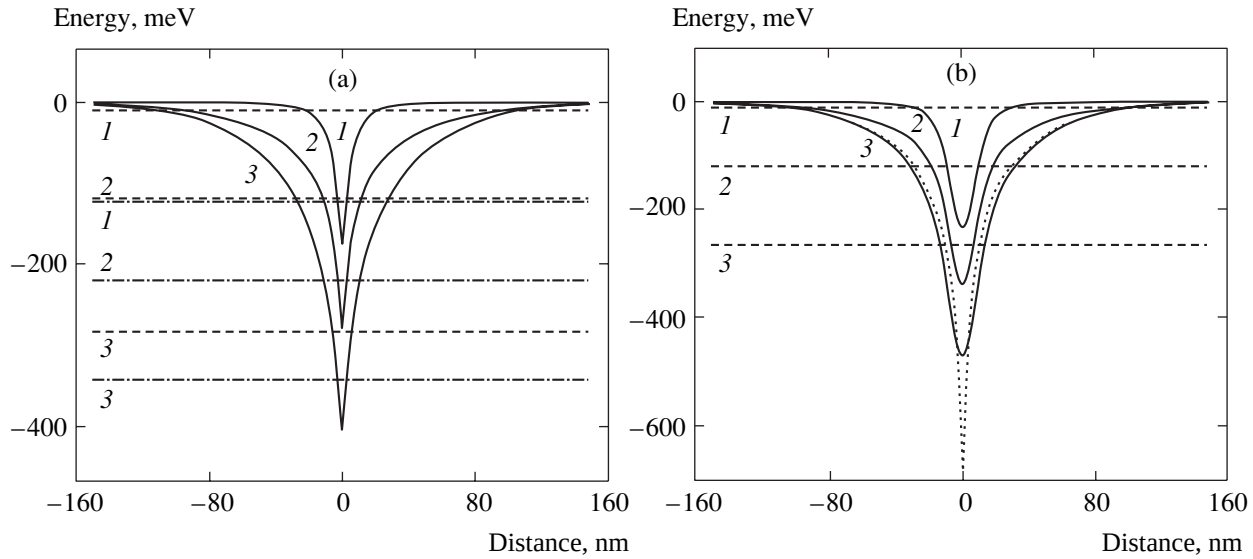

Fig. 6. The results of self-consistent calculations of the potential profiles (curves with minima), Fermi levels (dashed lines) and bottom of the lowest subband (dot-dash lines) for the electron temperatures 5 (*1*), 300 (*2*), and 700 K (*3*) in planar-doped GaAs:Si structures (a) with a donor concentration of $N_{Si}^{+} = 5 \times 10^{12}$ cm^{-2} and the Gaussian distribution of donors with a halfwidth of $d = 2$ nm and (b) with $N_{Si}^{+} = 1.5 \times 10^{13}$ cm^{-2} and $d = 20$ nm or (dashed profile) with the true δ-like distribution ($d = 0$), $T = 77$ K.

distances below 30 nm and 0.4 nm for the greater distances. The self-consistent calculations were continued until the potential at each of 600 points ceased to change to within 0.1% accuracy. The number of energy levels in the potential well was determined automatically from the condition that the maximum energy level differed from the zero-level energy only to within several percent of a preset electron temperature. The number of iterations necessary to achieve this accuracy varied from 30 to 150, depending on the temperature and electron density.

Figure 6 shows the results of our calculations for the Gaussian distributions with the halfwidths $d = 2$ nm (a) and 20 nm (b) and the donor concentrations $N_{Si}^{+} = 5 \times 10^{12}$ cm^{-2} (a) and 1.5×10^{13} cm^{-2} (b) and for the true δ-like distribution (dashed curve in Fig. 6b) for various temperatures. These results clearly demonstrate strong temperature dependence of the shape and depth of the potential well, which was reported in [11, 16]. This behavior of the potential has clear physical meaning: as the temperature increases, occupation of the higher sublevels and the corresponding increase in the average distance of the electron cloud from the center of positively charged layer unavoidably lead to an increase in the potential well depth. According to the results of our calculations, the depth increases almost linearly when the temperature increases from 5 to 700 K, at the same rate of 0.38 meV/K for both dopant concentrations.

A comparison of the shapes of the potential well for the δ-like and Gaussian distributions (Fig. 6b) shows that expansion of the doped region at a given dopant concentration leads to a decrease in the well depth, but only slightly (especially at high temperatures) changes is shape. It should be noted that our results obtained for low temperatures show very good agreement with published data [1, 22, 43].

3.3. Review and Analysis of Experimental Data

We believe [4, 16] that the observed decrease in the current is related to the electron trapping at metastable DX^{-} centers formed in strong electric fields in planar-doped GaAs:Si structures.[2] In order to estimate the concentration of DX^{-} centers formed in these structures, it is necessary to find a quantitative relationship between this decrease in the current and a change in the density of free electrons. Previously, it was established that the formation of trapping centers in homogeneously doped regions at weak electric fields is accompanied by an increase in the electron mobility. This increase was explained by a decrease in the number of scattering events due to the formation of neutral centers [47] or due to the correlation of positive and negative ions [48–53], which also leads to a decrease in the scattering probability.

In our control measurements with weak (several volt per centimeter) electric fields, a change in the current during the formation of trapping centers was significantly lower as compared to that observed in strong fields. We explain this behavior by an increase in the electron mobility in weak fields, which compensates a

[2] A decrease in the current in strong electric fields was previously observed in field-effect transistors, where the trapping centers were formed in AlGa–AlGaAs barriers [44–46], and also attributed to the DX^{-} centers.

change in the current related to a decrease in the number of electrons (in agreement with the results for homogeneously doped layers). However, the scattering on impurities in strong fields plays a rather insignificant role [54] and it is reasonable to suggest that a change in the current observed in our measurements involving hot electrons is determined primarily by changes in the density of free electrons.

Now we will separately consider the most important features of the phenomenon under consideration and compare the behavior observed with the results for homogeneously doped layers.

1. A decrease in the current in a strong electric field was observed only at low temperatures. The effect is qualitatively the same in the temperature range from 77 to 4.2 K.

2. It was established previously [4] and confirmed in this study that the new value of conductivity of a GaAs:Si structure remains unchanged for several days after switching off the electric field. However, heating the sample to a temperature above 120–140 K or exposure to a light with a wavelength below 0.9 μm restores the current on the initial level. It was found that such cycles (decreasing the current in a strong field at a low temperature followed by restoration on heating or illumination) could be multiply repeated without visible changes.

This behavior is similar to that observed for homogeneously doped $Al_xGa_{1-x}As$:Si layers with $x > 0.23$ (or with lower x and undoped GaAs at elevated pressures) under the conditions of DX⁻ center formation [26, 50, 55–59]. It is this analogy that allows us to identify the trapping centers formed in strong electric field with the DX⁻ centers [4].

3. A remarkable peculiarity of the observed effect is the partial restoration of current in the field lower than that used for the formation of trapping centers. The relative increase in the current depends on the applied voltage.

There are two basic differences between the trapping centers in homogeneously doped and planar-doped structures.

4. An important experimental fact is the existence of a well-defined threshold with respect to the donor concentration for the field-induced decrease in the current. No changes in the current were observed in a field of about 10 kV/cm in samples with a silicon concentration of 4×10^{12} cm^{-2}. This planar concentration corresponds to a bulk value of 8×10^{18} cm^{-3}. In homogeneously doped layers, metastable centers were observed at least for 2×10^{17} cm^{-3} [55, 60, 61]. To our knowledge, no data on such thresholds were reported for homogeneously doped $A^{III}B^{V}$ compounds or their alloys.

5. Another remarkable feature of the GaAs:Si structures studied is that a decrease in the current is proportional to the square root of time in the initial stage and can be described by a simple function (1) for large times.

Most of the previous investigations of DX⁻ centers in homogeneously doped $A^{III}B^{V}$ compounds were performed under the conditions (hydrostatic pressure, alloys with large x) where the main state of silicon corresponded to DX⁻ centers with an energy level in the forbidden band. In this case, the energy barrier for the electron emission from DX⁻ centers is higher than the barrier for electron trapping. Therefore, the occupancy of DX⁻ centers in the state corresponding to thermal equilibrium (the degree of which depends on x or the pressure) becomes measurable only upon cooling the sample below a certain temperature (120–140 K). At this temperature, on the one hand, a time for the over-barrier trapping of free electrons from the Γ band is small as compared to the time of measurements and, on the other hand, the probability of ejection of the trapped electron is negligibly small. However, the results of theoretical calculations [62–65] show that the DX⁻ center in GaAs at atmospheric pressure corresponds to a high resonance metastable state in the conduction band, having an effective energy of about 0.230–0.295 eV relative to the Γ band edge, which increases with the dopant concentration [26, 47, 66]. The barrier for the electron ejection from this state is about 0.30–0.33 eV, while the barrier for trapping electrons with energies at the conduction band bottom is about 0.55–0.63 eV.

A significant (measurable) population of these centers in thermal equilibrium is possible only provided that the conduction band is filled with electrons approximately up to the level of the DX⁻ center. Therefore, the electron density must be about 1.5×10^{19} cm^{-3}. Probably, the difference between the density of free electrons and the dopant concentration in heavily homogeneously doped samples is related (at least partly) to the presence of DX⁻ centers [8] (this was already mentioned in Section 3.1). Generally speaking, it is theoretically possible to form a DX⁻ center in GaAs even at lower electron densities. For example, this can be achieved by the optical excitation of free or valence electrons into a resonance state of silicon atom with a zero generalized coordinate Q, followed by its transformation into the DX⁻ center ($Q \neq 0$) due to a strong electron–phonon interaction in this state (see [67] and references therein) or due to the Jahn–Teller interaction [68]. However, the resonance character of this interaction and, hence, the short lifetime make the electron relaxation within the conduction band more probable than the trapping of a second electron with "large relaxation" of the lattice (corresponding to the DX⁻ center formation).

3.4. Formation of DX⁻ Centers during Redistribution of Hot Electrons in a Quantum Well

The results of calculations of the temperature dependence of the energy spectrum (Fig. 6) were used for describing the effect of a strong electric field on the shape of the potential in the quantum well formed as

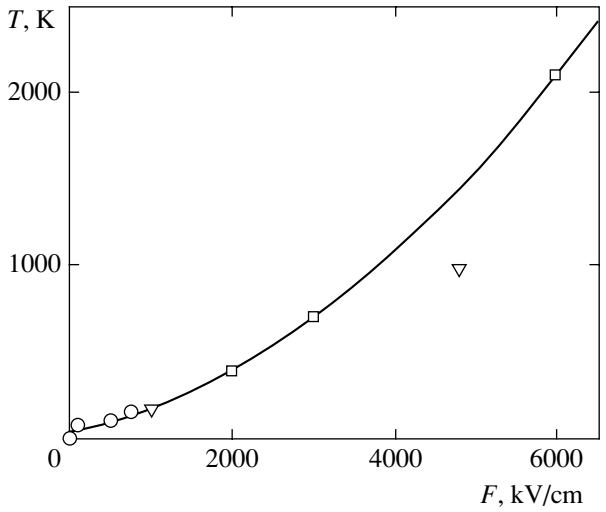

Fig. 7. Plot of the calculated electron temperature versus electric field strength for δ-doped GaAs with an impurity concentration $N = 5 \times 10^{11}$ cm^{-2} at a lattice temperature T = 4 K [69] (circles) and $N = 4 \times 10^{11}$ cm^{-2} at T = 5 K [70] (triangles) and for a homogeneously doped GaAs [71] without scattering on impurities (squares).

a result of planar doping. The self-consistent potential is independent of the lattice temperature: the temperature in the Fermi function refers to the electron temperature. At high electron densities characteristic of our samples, the electron–electron interactions predominate over other inelastic scattering processes even in strong electric fields. Therefore, the electron distribution function can be approximated by the Fermi distribution with an effective electron temperature. We did not study the relationship between the electron temperature and the electric field strength, which can be determined using the Monte Carlo method or by numerical solution of the Boltzmann equation. However, in order to evaluate the order of magnitude, we extrapolated the results of calculations [69, 70] performed for higher field strengths and took into account some results for the homogeneously doped GaAs [71]. As can be seen from Fig. 7, a temperature of about 700 K corresponds to a field of 3 kV/cm, that is, to the observed threshold for DX$^-$ center formation.

In order to explain the mechanism of DC$^-$ center formation in a strong electric field, it is important to take into account that the energy level of DX$^-$ centers in the absence of such fields occurs at $E_{DX} > 0$ in the structures of both types at T below approximately 200 K (see Fig. 6). Fore this reason, the wave functions of quasi-free electrons and the electron state corresponding to the DX$^-$ center overlap insignificantly and the probability of transitions between these sates is small. In a strong electron field, the energy level of the DX$^-$ centers shifts to fall within the potential well and the formation of DX$^-$ centers becomes principally possible.

In addition, the electric field increases the screening radius [16]. According to [62], an increase in the screening radius leads to a decrease in the energy of DX$^-$ centers and favors the shift of Si_{Ga} toward the interstitial position.

It should be emphasized that, as was demonstrated previously [4], a decrease in the current depends only on the product of the pulse duration and the frequency (within the interval of pulse durations studied). This implies that an increase in the lattice temperature during the pulse does not influence the DX$^-$ center formation. Therefore, the system is not in thermodynamic equilibrium and, strictly speaking, the dynamics of DX$^-$ center formation proposed in [62] is inapplicable.

However, is necessary to take into account that hot carriers may excite, in addition to optical and acoustic phonons, the local oscillations of silicon atoms.[3] Such a local model can decay into phonons of lower energies during a short period of time [72]. If the probability of binding two hot carriers on a substitutional silicon impurity is comparable with the probability of decay of the local vibrational modes, the formation of DX$^-$ centers also becomes possible. Apparently, this process must significantly depend on the electron density, which probably explains the experimentally observed threshold with respect to the donor concentration.

The processes in planar-doped structures are characterized by the hierarchy of the electron trapping times, which is related to a multilevel structure of the energy spectrum and the potential fluctuations in the plane of doping [73, 74]. The structure under consideration resembles the systems scale-invariant with respect to time [75] and its behavior is descried by a "delayed" exponent observed in our experiments.

4. CONCLUSIONS

We have thoroughly studied the kinetics of current changes in planar-doped GaAs:Si structures in strong electric fields at low temperatures (T = 77 and 4.2 K). The obtained data, together with the results of our previous investigations, show that (i) a decrease in the current of hot electrons is caused by the formation of DX$^-$ centers in the plane of doping and (ii) this process is characterized by an important feature distinguishing it from the DX$^-$ center formation in homogeneously doped crystals: the formation of DX$^-$ centers in the GaAs:Si structures studied is characterized by thresholds with respect to the electric field strength and the electron density (or donor concentration).

In order to explain this behavior, we calculated the depth and shape of the potential well and the position of the Fermi level for various electron temperatures. As is

[3] To our knowledge, no theoretical investigations or experimental observations of local oscillations excited by hot electrons have been reported. However, this possibility may be indicated by an additional phonon flux observed in the time-of-flight spectra [9].

known, the potential well in δ-doped structures is formed due to the positive charge of ionized impurity and the space charge of electrons. In a strong field, the cloud of hot electrons shifts to a greater distance from the plane of doping and the potential well increases in both width and depth.

In a zero electric field, the energy level of DX^- centers is significantly above the well edge and there is virtually no electrons with such energies in the well, which makes the probability of DX^- center formation negligibly small. In a strong field, the electron temperature increases, the edge of the well raises, and the energy level of DX^- centers falls within the well, which makes the formation of DX^- centers possible. This explains the existence of a threshold with respect to the electric field strength. The estimates of the electron temperature obtained from the results of our measurements give a threshold value close to that observed in experiments. The existence of a threshold with respect to the electron density can be explained within the framework of the hypothesis about the formation of DX^- centers involving the excitation of local vibrational modes by hot electrons.

Interpretation of the observed kinetics of DX^- center formation in planar-doped GaAs:Si, having the shape of a "delayed" exponent, requires further development of the proposed model.

REFERENCES

1. A. Zrenner, F. Koch, R. L. Williams, *et al.*, Semicond. Sci. Technol. **3**, 1203 (1988).
2. P. M. Koenraad, W. de Lange, F. A. P. Blom, *et al.*, Semicond. Sci. Technol. **6**, B143 (1991).
3. E. Bucks, M. Heilblum, Y. Levinson, and H. Striktman, Semicond. Sci. Technol. **9**, 2031 (1994).
4. R. Stasch, M. Asche, L. Daeweritz, *et al.*, J. Appl. Phys. **77**, 4463 (1995).
5. R. Stasch, M. Asche, M. Giehler, *et al.*, J. Phys.: Condens. Matter **8**, 3215 (1996).
6. T. M. Schmidt and A. Fazzio, Phys. Rev. B **51**, 7898 (1995).
7. A. Fazzio and T. M. Schmidt, Mater. Sci. Forum **196**–**201**, 421 (1995).
8. R. H. Miwa and T. M. Schmidt, Appl. Phys. Lett. **74**, 1999 (1999).
9. B. Danilchenko, A. Klimashov, M. Asche, and R. Hey, Semicond. Sci. Technol. **15**, 744 (2000).
10. M. Asche, O. G. Sarbey, and Z. Wilamowski, Physica B (Amsterdam) **272**, 241 (1999).
11. M. Asche and O. G. Sarbey, Phys. Rev. B **65**, 233309 (2002).
12. L. Daeweritz, H. Kostial, R. Hey, *et al.*, J. Cryst. Growth **150**, 214 (1955).
13. D. K. Biegelsen, R. D. Bringans, J. E. Northrup, and L.-E. Swartz, Phys. Rev. B **41**, 5701 (1990).
14. Z. M. Wang, L. Daeweritz, P. Schuetzenduebe, and K. H. Ploog, Phys. Rev. B **61**, R2440 (2000).
15. R. Stasch, private communication.
16. M. Asche and O. G. Sarbey, Physica B (Amsterdam) **308**–**310**, 788 (2001).
17. J. Wagner, M. Ramsteiner, W. Stolz, *et al.*, Appl. Phys. Lett. **55**, 978 (1989).
18. O. Brandt, G. E. Crook, K. Ploog, *et al.*, Appl. Phys. Lett. **59**, 2730 (1991).
19. P. M. Koenraad, F. A. P. Blom, C. J. G. M. Langerak, *et al.*, Semicond. Sci. Technol. **5**, 861 (1990).
20. K. Koehler, P. Ganser, and M. Maier, J. Cryst. Growth **127**, 720 (1993).
21. M. R. Fahy, M. J. Ashwin, J. J. Harris, *et al.*, Appl. Phys. Lett. **61**, 1805 (1992).
22. M. J. Ashwin, M. Fahy, J. J. Harris, *et al.*, J. Appl. Phys. **73**, 633 (1993).
23. R. C. Newman, E. G. Grosche, M. J. Ashwin, *et al.*, Mater. Sci. Forum **258**–**263**, 1 (1997).
24. M. J. Ashwin, R. C. Newman, and K. Muraki, J. Appl. Phys. **82**, 137 (1997).
25. R. Murray, R. C. Newman, M. J. L. Sangster, *et al.*, J. Appl. Phys. **66**, 2589 (1989).
26. T. N. Theis, M. Mooney, and S. L. Wright, Phys. Rev. Lett. **60**, 361 (1988).
27. E. Skuras, R. Kumar, R. L. Williams, *et al.*, Semicond. Sci. Technol. **6**, 535 (1991).
28. S. Arscott, M. Missous, and L. Dobaczewski, Semicond. Sci. Technol. **7**, 620 (1992).
29. J. A. Wolk, M. B. Krueger, J. N. Heyman, *et al.*, Phys. Rev. Lett. **66**, 774 (1991).
30. M. Saito, A. Oshiyama, and O. Sugino, Phys. Rev. B **47**, 13205 (1993).
31. R. Jones and S. Oeberg, Phys. Rev. B **44**, 3407 (1991).
32. G. Li and C. Jagadish, Appl. Phys. Lett. **70**, 90 (1997).
33. L. Pauling and Z. S. Herman, Phys. Rev. B **28**, 6154 (1983).
34. D. H. Rich, T. Miller, and T. C. Chiang, Phys. Rev. B **37**, 3124 (1988).
35. E. Artacho and F. Yndurain, Phys. Rev. B **42**, 11310 (1990).
36. J. Dabrowski, F. Pehlke, and M. Scheffler, Phys. Rev. B **49**, 4790 (1994).
37. K. Shiraishi and T. Ito, Jpn. J. Appl. Phys., Part 2 **37**, L1211 (1998).
38. R. B. Beall, J. B. Clegg, J. Castagne, *et al.*, Semicond. Sci. Technol. **4**, 1171 (1989).
39. C. R. Proetto, in *Delta-Doping of Semiconductors*, Ed. by E. F. Schubert (Cambridge Univ. Press, Cambridge, 1996), p. 23.
40. A. Zrenner, F. Koch, and K. Ploog, Surf. Sci. **196**, 671 (1988).
41. J. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
42. Uday Gupta and A. K. Rajagopal, Phys. Rev. A **22**, 2792 (1980).

43. O. A. Mezrin, A. Y. Shik, and V. O. Mezrin, Semicond. Sci. Technol. **7**, 664 (1992).
44. R. Fischer, T. J. Drummond, J. Klem, *et al.*, IEEE Trans. Electron Devices **31**, 1628 (1984).
45. A. Kastalsky and R. A. Kiehl, IEEE Trans. Electron Devices **33**, 414 (1986).
46. T. N. Theis, B. D. Parker, P. M. Solomon, and S. L. Wright, Appl. Phys. Lett. **49**, 1542 (1986).
47. D. K. Maude, J. C. Portal, L. Dmowski, *et al.*, Phys. Rev. Lett. **59**, 815 (1987).
48. D. J. Chadi, K. J. Chang, and W. Walukiewisz, Phys. Rev. Lett. **62**, 1923 (1989).
49. E. P. O'Reilly, Appl. Phys. Lett. **55**, 1409 (1989).
50. T. Suski, R. Piotrzkowski, P. Wisniewsky, *et al.*, Phys. Rev. B **40**, 4012 (1989).
51. J. Kossut, Z. Wilamowski, T. Dietl, and K. Swiatek, in *Proceedings of 20th ICPS*, Ed. by E. M. Anastassaki and J. D. Joanopoulos (World Sci., Singapore, 1990), p. 613.
52. P. Wisniewski, T. Suski, L. H. Dmowski, *et al.*, Mater. Sci. Forum **143**–**147**, 617 (1994).
53. E. Litwin-Staszewska, T. Suski, C. Skierbiszewski, *et al.*, J. Appl. Phys. **77**, 405 (1995).
54. W. T. Masselink, T. S. Henderson, J. Klem, *et al.*, IEEE Trans. Electron Devices **33**, 639 (1986).
55. M. Tachikawa, T. Fujizawa, H. Kukimoto, *et al.*, Jpn. J. Appl. Phys. **24**, L893 (1985).
56. T. Ishikawa and T. Yamamoto, Jpn. J. Appl. Phys. **25**, L484 (1986).
57. J. T. Fujisawa and H. Kukimoto, Jpn. J. Appl. Phys. **29**, L388 (1990).
58. E. Calleja, F. Garcia, A. Gomez, *et al.*, Appl. Phys. Lett. **56**, 934 (1990).
59. M. F. Li and P. Y. Yu, Jpn. J. Appl. Phys., Suppl. **32**, 200 (1993).
60. M. Mizuta, M. Tachikawa, H. Kukimoto, and S. Minomura, Jpn. J. Appl. Phys. **24**, L143 (1985).
61. M. F. Li, P. Y. Yu, E. R. Weber, and W. Hansen, Appl. Phys. Lett. **51**, 349 (1987).
62. D. J. Chadi and K. J. Chang, Phys. Rev. B **39**, 10063 (1989).
63. S. W. Biernacki, Semicond. Sci. Technol. **11**, 17 (1996).
64. T. M. Schmidt, A. Fazzio, and M. J. Caldas, Phys. Rev. B **53**, 1315 (1996).
65. D. J. Chadi and K. J. Chang, Phys. Rev. Lett. **61**, 873 (1988).
66. T. Baba, M. Mizuta, T. Fujisawa, *et al.*, Jpn. J. Appl. Phys. **28**, L891 (1987).
67. Y. Toyozawa, Physica B (Amsterdam) **116**, 7 (1983).
68. P. Krueger and J. Polman, Phys. Rev. Lett. **74**, 1155 (1995).
69. S. Goodnik and P. Lugli, in *High-Speed Electronics*, Ed. by B. Kallack and H. Beneking (World Sci., Springer, 1986), p. 116.
70. P. Lugli, S. Goodnick, and F. Koch, Superlattices Microstruct. **2**, 335 (1986).
71. J. Pozela and A. Reklaitis, Solid-State Electron. **23**, 927 (1980).
72. C. Newman, in *Delta-Doping of Semiconductors*, Ed. by E. F. Schubert (Cambridge Univ. Press, Cambridge, 1996), p. 279.
73. Z. Wilamowski, J. Kossut, W. Jantsch, and G. Ostermayer, Semicond. Sci. Technol. **6**, B38 (1991).
74. P. Sobkowisz, Z. Wilamowski, and J. Kossut, Semicond. Sci. Technol. **7**, 1155 (1992).
75. H. Scher, M. F. Shlesinger, and J. F. Bendler, Phys. Today **44**, 26 (1991).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 3, 2004, pp. 585–601.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 3, 2004, pp. 672–690.

SOLIDS
Electronic Properties

The Pseudogap Mode as Long-Lived States of Noncoherent Pairs with Large Momenta

V. I. Belyavskiĭ*, Yu. V. Kopaev, Yu. N. Togushova, and S. V. Shevtsov**
Lebedev Physical Institute, Russian Academy of Sciences, Leninskiĭ pr. 53, Moscow, 119991 Russia
**e-mail: vib@spu.ac.ru*
***e-mail: kopaev@lebedev.sci.ru*
Received January 19, 2004

Abstract—The mirror nesting of the Fermi contour of a quasi-two-dimensional electronic system and the presence of at least one negative eigenvalue of the Fourier transform of interaction energy are sufficient conditions for the formation of bound states of the relative motion of pairs with large total momenta. As distinct from pairing by attractive interactions, the wave functions of such pairs have alternating signs and lines of zeros that twice intersect the Fermi contour in the regions of definition of relative motion momenta. The total number of intersection points between the line of zeros and the Fermi contour is determined by symmetry of a linear combination of the wave functions of crystallographically equivalent pairs. Long-lived quasi-stationary states exist in the form of noncoherent pairs with different but close momenta and cause substantial suppression of the density of one-particle states (the appearance of a pseudogap) over a fairly wide energy range. The upper temperature bound of the pseudogap is determined by the decay of pairs, and the lower bound, by phase coherence disturbance when pairs leave the condensate that is formed at some optimum pair momentum value owing to mirror nesting. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The attraction of particles excited outside (or holes excited inside) the Fermi surface results in the formation of a true bound state of a pair with zero total momentum [1] (Cooper pair) and decreases the energy of the system. This is evidence of instability of the ground state [2], which is treated at zero temperature as a fully occupied momentum space region within the Fermi surface. The energy of the bound state of a Cooper pair exponentially depends on the dimensionless parameter, the product of the effective coupling constant and the density of states on the Fermi surface. The preexponential factor is determined by the energy width of the layer adjacent to the Fermi surface in which attraction is defined that prevails over Coulomb repulsion. The separation of a layer with predominant attraction means the dynamic bounding of the phase space region that contributes to relative pair motion states. For pairs with zero total momentum, there are no kinematic constraints, and the momenta of particles coincide with the momenta of relative motion and can be arbitrary. The Fermi surface is the natural boundary that separates momentum space regions where pairs of particles (Brillouin zone region external with respect to the Fermi surface) and pairs of holes (internal Fermi surface region) can be excited. At this boundary, the energy of pair excitation vanishes. Note that, in addition to the stationary state with a negative energy (a discrete level splitting off from the continuous spectrum band), the Cooper problem of two attracting particles formally admits solutions that belong to the continuous spectrum band. The weight of each such state of the relative motion of a Cooper pair is, however, small (inversely proportional to the number of particles) compared with the weight of the stationary state (which is on the order of magnitude of one). We can therefore ignore the redistribution of levels in the continuous spectrum band.

For pairs with nonzero total momenta, the momentum space regions accessible to the pairing of both electrons and holes are substantially reduced [3–5]. This kinematic constraint decreases the statistical weight of a pair (of the number of quantum states that contribute to the wave function of the relative motion of the pair) and, accordingly, the bound state energy. The locus corresponding to zero excitation energy of the pair is then a set of a smaller dimension than a surface in the three-dimensional momentum space, as for pairs with zero momenta. For this reason, pairing becomes possible at a not arbitrarily small coupling constant. The superconducting pairing channel with zero momentum therefore stands out as predominant [6].

Cuprate superconducting compounds with layered crystal structures, which are quasi-two-dimensional ($2D$) electronic systems, exhibit certain properties (in both the superconducting and normal states) that do not fit in [7] with the Bardin–Cooper–Schriffer theory of superconductivity. Most clearly, the difference between usual and high-temperature cuprate superconductors manifests itself by the pseudogap behavior of the latter at a relatively low hole doping level in a fairly wide temperature interval above the superconducting transi-

tion temperature [8]. This behavior corresponds to substantial suppression of the density of one-particle states. It can naturally be explained on the assumption that pairs of carriers whose condensation is responsible for superconductivity dissociate not at the condensation temperature, at which phase coherence of pairs disappears, but at a much higher temperature that corresponds to pseudogap opening. In usual superconductors, the region of fluctuations close to the superconducting transition temperature is extremely narrow, which is responsible for the short lifetime of pairs outside the condensate. It follows that, in the Bardin–Cooper–Schriffer scheme corresponding to the attraction of carriers that form pairs, the transition to the noncoherent state is almost immediately accompanied by unpairing. This circumstance provides grounds for claiming that the Bardin–Cooper–Schriffer scheme based on the mean-field approximation is inapplicable to the problem of high-temperature superconductivity of cuprates.

As in the Bardin–Cooper–Schriffer model, the superconducting state of cuprates arises because of singlet pairing of carriers in copper–oxygen planes. The electronic 2*D* spectrum of cuprates, however, differs from that of the isotropic Bardin–Cooper–Schriffer model. The line that separates the occupied and unoccupied states in the 2*D* momentum space, the Fermi contour, is in its essential part situated in the neighborhood of the saddle point of the electronic dispersion law at doping levels close to half-doping [9]. The characteristic shape of the Fermi contour in cuprates with hole doping is evidence for strong anisotropy of the Fermi velocity and effective masses of electrons and holes, which strengthens the logarithmic Van Hove singularity of the 2*D* density of states. For this reason, the principal values of the tensor of reciprocal effective masses, which determine the kinetic energy of the relative motion of a pair, have different signs in a broad neighborhood of the saddle point.

The special features of the dispersion law in the vicinity of the saddle point manifest themselves when particles with a large total momentum (on the order of twice the Fermi momentum) are paired [5]. The energy of such a pair is the sum of the energy of the center of mass and the energy of relative motion. The bound state, if it appears, corresponds to a decrease in the energy of relative motion. Because of the kinematic constraints on pairing with a nonzero pair momentum in the 2*D* system, the boundary that separates the regions of filled and vacant states is generally formed by two points. For this reason, the bound state of a pair can only arise on the condition that the effective coupling constant exceeds some threshold value, because the density of states of relative motion vanishes at zero pair excitation energy.

If, for some distinguished directions and total pair momentum values, the electronic dispersion law admits the formation of a boundary between filled and vacant states in the form of segments of lines that are Fermi contour regions, these segments constitute the pair Fermi contour, which plays the same role for relative pair motion as the Fermi contour for the motion of particles [10, 11]. The appearance of a pair Fermi contour is possible provided the mirror nesting condition is satisfied, that is, when separate Fermi contour segments coincide with the isoline of the kinetic energy of relative pair motion [12]. This condition can be exactly satisfied when the pair momentum is oriented along one of the symmetrical directions in the 2*D* Brillouin zone and the Fermi contour has a special form that determines the pair momentum value. Perfect mirror nesting is a necessary condition for the formation of a bound state of the relative motion of a pair with a nonzero total momentum at an arbitrary (arbitrarily small) effective coupling constant [13]. Bound states are also retained (with lower bound state energies) when mirror nesting is approximate, and they disappear at some characteristic deviation from perfect nesting [12].

Currently, the nature of pairing interactions in cuprates cannot be considered established unambiguously. In particular, along with usual electron–phonon pairing attraction [14, 15], pairing mechanisms caused by repulsive interaction are discussed [16]. For this reason, the solution to the problem of two repelling particles constituting a pair with a large total momentum can be useful for determining the nature of the superconducting state of cuprates. The simplest approximation corresponding to point repulsive interaction (as in the Bardin–Cooper–Schriffer scheme) shows [11] that two solutions to the Cooper problem arise in a continuous spectrum band of the relative motion of a pair with a large total momentum. These solutions are characterized by complex energies whose imaginary parts have different signs. One of these solutions with the smaller real energy part has negative decay, which can result in a tendency toward system instability with respect to repulsive interaction. The second solution with the correct decay sign corresponds to the redistribution of levels in the continuous spectrum band. It can, in particular, be responsible for the formation of a quasi-stationary state with a fairly long lifetime [11].

According to commonly accepted views [17], the physics of cuprates is determined by the very strong repulsive interaction of electrons (or holes) localized on copper atoms in the conducting copper–oxygen planes. It is believed that the Hubbard model, which only includes correlations on one and the same node, and more complex related models that also include fairly well-screened internode interactions (the Penson–Colb–Hubbard model [18]) are well suited for describing strong repulsive electron–electron correlations. The higher the Hubbard energy of intracenter correlation U_0 compared with the energy bandwidth t ($U_0 \gg t$), the higher the predictive ability of such models. It appears that approximately equal U_0 and t values correspond to real cuprate compounds. The physics of cuprates can therefore be qualitatively studied in terms of band mod-

els, which give the best results when the opposite inequality $t \gg U_0$ holds. In the present work we use this approach taking into account the metric of the momentum space in the region of momentum definition for relative pair motion to solve the problem of bound states of pairs of likely charged particles with large total momenta and repulsive interaction in a quasi-two-dimensional electronic system.

2. THE KINEMATICS OF PAIRS

Let $\varepsilon(\mathbf{k})$ be the law of dispersion of particles in a 2*D* electronic system. The $\varepsilon(\mathbf{k}) = \mu$ isoline, where μ is the chemical potential, then determines the Fermi contour, and a pair of likely charged particles with the total momentum $\mathbf{K}$ and the energy of the center of mass $\varepsilon_{cm}(\mathbf{K}) = 2\varepsilon(\mathbf{K}/2)$ should have momenta $\mathbf{k}_+$ and $\mathbf{k}_-$ either inside $[\varepsilon(\mathbf{k}_\pm) < \mu]$ or outside $[\varepsilon(\mathbf{k}_\pm) > \mu]$ the Fermi contour, as is schematically shown in Fig. 1. The momenta of the particles constituting the pair, like the relative pair motion momentum $\mathbf{k} = (\mathbf{k}_+ - \mathbf{k}_-)/2$, are within some Ξ_K region of the momentum space, which, generally speaking, consists of two subregions, namely, the filled (at $T = 0$) $\Xi_K^{(-)}$ and vacant $\Xi_K^{(+)}$ regions. It follows that the momenta of particles excited above the Fermi contour belong to $\Xi_K^{(+)}$, and the momenta of holes excited inside the Fermi contour, to $\Xi_K^{(-)}$. The Ξ_K region, like its subregions $\Xi_K^{(-)}$ and $\Xi_K^{(+)}$, is symmetrical with respect to the inversion of the relative pair motion momentum, $\mathbf{k} \longrightarrow -\mathbf{k}$. Note that hole and electronic fillings correspond to the $\Xi_K^{(-)}$ and $\Xi_K^{(+)}$ regions, respectively, in cuprates with high-temperature superconductivity.

The energy of a pair is the sum of the energy of the center of mass and the relative motion energy,

$$\varepsilon_K^{(r)}(\mathbf{k}) = \varepsilon(\mathbf{K}/2 + \mathbf{k}) + \varepsilon(\mathbf{K}/2 - \mathbf{k}) - 2\varepsilon(\mathbf{K}/2). \quad (1)$$

The metric of the 2*D* momentum space of relative pair motion (the metric of the momentum space in the vicinity of the $\mathbf{K}/2$ point) can be determined by representing (1) as a quadratic form corresponding to the first nonvanishing terms of the expansion of (1) in powers of momentum components of relative pair motion,

$$\varepsilon_K^{(r)}(\mathbf{k}) \approx D_{\alpha\beta}(\mathbf{K})k_\alpha k_\beta. \quad (2)$$

Here, the 2*D*-tensor $D_{\alpha\beta}(\mathbf{K})$ determines the reciprocal reduced masses, and the summation over repeating Greek indices from 1 to 2 is implied. If the principal values of this tensor have like signs, the metric of the momentum space of relative pair motion can be considered elliptical, because the kinetic energy isolines of relative motion are ellipses at energies close to zero (Fig. 1a). If the principal values of the $D_{\alpha\beta}(\mathbf{K})$ tensor have different signs, quadratic form (2) determines a hyperbolic metric, because isolines are then two families of hyperbolas in a narrow neighborhood of the $\mathbf{K}/2$ point. The structure of isolines over the whole Ξ_K region, which constitutes an extended neighborhood of the $\mathbf{K}/2$ point, is then more complex.

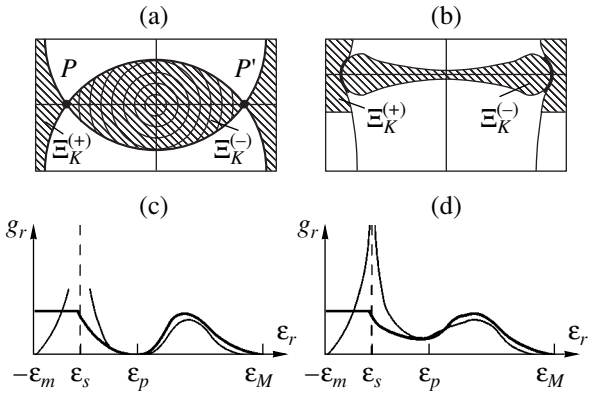

Fig. 1. Kinematically allowed momentum space region (hatched) for relative pair motion with the total momentum $\mathbf{K}$: (a) boundary separating the occupied ($\Xi^{(-)}$) and unoccupied ($\Xi^{(+)}$) region Ξ_K subregions degenerates into points *P* and *P*' and (b) boundary between $\Xi^{(-)}$ and $\Xi^{(+)}$ is segments (shown by tick lines) that form the pair Fermi contour. The $\mathbf{K}/2$ vector corresponds to the center of the Ξ_K region. The density of states of relative pair motion g_r as a function of the kinetic energy of relative motion: (c) pair Fourier contour degenerates into points [*P* and *P*' in (a)], and (d) pair Fourier contour of finite length [see (b)].

The density of relative motion states when the metric of the momentum space is elliptical is schematically shown by a thick line in Fig. 1c. When the boundary separating $\Xi_K^{(-)}$ and $\Xi_K^{(+)}$ consists of points (*P* and *P*' in Fig. 1a), zero density of states corresponds to the energies of the relative motion of the pair at these points, $\varepsilon_p = 2[\mu - \varepsilon(\mathbf{K}/2)]$. The smallness of the density of states in the vicinity of the ε_p energy corresponds to only taking into account the contribution of isoline regions that belong to the Ξ_K region (such isolines are shown for the $\Xi_K^{(-)}$ region in Fig. 1a). If the boundary between $\Xi_K^{(-)}$ and $\Xi_K^{(+)}$ was a line (the pair Fermi contour shown in Fig. 1b by a thick line), the density of states of the relative motion of the pair, g_r, at the elliptical metric of the Ξ_K region (shown by a thick line in Fig. 1d) would be finite at the ε_p energy. Continuous spectrum band edges correspond to the $-\varepsilon_m$ and ε_M relative motion energies ε_r. Clearly, $\varepsilon_m = 0$ if the metric is elliptical.

If the metric of the Ξ_K region is hyperbolic, the energy of relative pair motion can take both positive and negative values. The logarithmic Van Hove singularity of the density of states (thin lines in Figs. 1c and 1d) then corresponds to zero energy ($\varepsilon_s = 0$). The den-

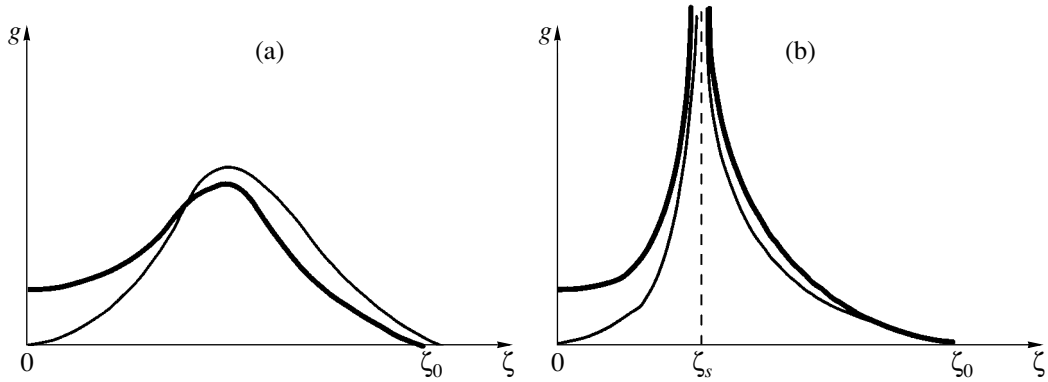

Fig. 2. Density of states for excited (a) pair of particles and (b) pair of holes. Thin lines are the densities of state when the pair Fermi contour degenerates into points (*P* and *P*' in Fig. 1a). The formation of the pair Fermi contour of a finite length results in a nonzero density of states at excitation energy $\zeta = 0$ (thick lines). Figure 2b corresponds to a hyperbolic metric of the momentum space in the $\Xi^{(-)}$ region; excitation energy ζ_s corresponds to the Van Hove 2*D* singularity. Figure 2a corresponds to an elliptical momentum space metric; the density of states for the excitation of pairs of both particles and holes then does not have singularities within the continuous spectrum band.

sity of states vanishes at the energy equal to ε_p (Fig. 1c) if the boundary between $\Xi_K^{(-)}$ and $\Xi_K^{(+)}$ degenerates into points. The appearance of the pair Fermi contour results in a finite value of the density of states at the ε_p point, as is shown in Fig. 1d.

The excitation energy ζ of a pair of particles outside the region bounded by the Fermi contour corresponds to the kinetic energy of relative pair motion $\varepsilon_K^{(r)}(\mathbf{K}) \geq \varepsilon_p$, and the density of states of such excitations over this energy range coincides with the density of states of relative pair motion $g_r(\varepsilon_K^{(r)})$, as is shown in Fig. 2a. If a pair of holes is excited inside the region bounded by the Fermi contour, the excitation energy (positive by definition) corresponds to the kinetic energy of the relative motion of a pair of particles in the interval $0 \leq \varepsilon_K^{(r)}(\mathbf{k}) \leq \varepsilon_p$ for the elliptical metric and $-\varepsilon_m \leq \varepsilon_K^{(r)}(\mathbf{k}) \leq \varepsilon_p$ for the hyperbolic metric. The density of states of such excitations that corresponds to these intervals of the kinetic energy of relative pair motion reproduces the density of states of relative motion g_r shown in Fig. 1 up to a change in the origin for counting energy and mirror reflection at the ε_p point, as is shown for the hyperbolic metric in Fig. 2b.

At zero excitation energy ($\zeta = 0$), the densities of excitation states of a pair of particles and a pair of holes coincide. If the momentum space metric is hyperbolic, the behaviors of the densities of states at $\zeta > 0$ are essentially different for pairs of particles [$g(\zeta)$ is a regular function without singularities over the whole range $0 \leq \zeta \leq \zeta_0$] and pairs of holes [$g(\zeta)$ exhibits a logarithmic singularity at the $\zeta_s = 2\mu - 2\varepsilon(\mathbf{K}/2)$ point, which can be close to the left edge of the continuous spectrum band]. It follows that, if ζ_s is small, and it is small under the conditions of a strong anisotropy of the effective masses in an extended neighborhood of the saddle point, electron–hole asymmetry of the density of states of relative motion is well defined. There is no such asymmetry for the elliptical metric, when the functions representing the density of states of the relative motion of a pair of electrons and a pair of holes are qualitatively similar; that is, the density of excitation states of a pair of particles and a pair of holes has the form shown in Fig. 2a.

The appearance of the pair Fermi contour, which results in a nonzero density of states of relative motion at $\zeta = 0$, is an obvious condition necessary for pairing at $K \neq 0$ to be possible at an arbitrarily small effective coupling constant. By definition, the pair Fermi contour is the 2*D* momentum space locus where Fermi contour segments coincide with the kinetic energy isolines of relative pair motion corresponding to zero excitation energy, $\varepsilon_K^{(r)}(\mathbf{K}) = \varepsilon_p$. As the one-particle states on the pair Fermi contour with the momenta $\mathbf{K}/2 \pm \mathbf{k}$, which correspond to a pair with the total momentum $\mathbf{K}$ and the relative motion momentum $\mathbf{k}$, have energies equal to the Fermi energy μ, the equations that determine the pair Fermi contour can be written as $\varepsilon(\mathbf{K}/2 \pm \mathbf{k}) = \mu$. These equations determine two lines in the 2*D* momentum space of the relative motion of a pair with the total momentum $\mathbf{K}$. These are the coinciding segments of these two lines that form the pair Fermi contour. The region of relative motion momenta bounded by these lines, for which the pair Fermi contour is part of its external boundary, is the $\Xi_K^{(-)}$ region defined above. The region external with respect to $\Xi_K^{(-)}$, which also has the pair Fermi contour as part of its boundary (the boundary of this region includes segments of the 2*D* Brillouin zone boundary) is the $\Xi_K^{(+)}$ region. The character of the solutions to the system of equations $\varepsilon(\mathbf{K}/2 \pm$

k) = μ is shown in Figs. 3a and 3b for the cases when the pair Fermi contour degenerates into two points and when it has a finite length, respectively. In Fig. 3a, which corresponds to the isotropic dispersion law, k_F is the Fermi momentum and $k_p = \sqrt{k_F^2 - K^2/4}$; the structure of the relative motion energy isolines is shown in the $\Xi_K^{(-)}$ region. In Fig. 3b, where the Fermi contour is a square with rounded corners, the $\Xi_K^{(-)}$ region is hatched.

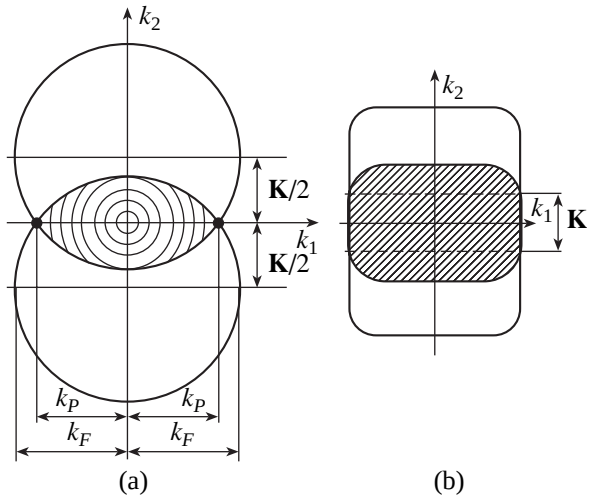

Fig. 3. Character of solutions to the system ε(**K**/2 ± **k**) = μ when the pair Fermi contour (a) degenerates into two points and (b) has a finite length; the $\Xi_K^{(-)}$ region is hatched.

3. ELEMENTARY EXCITATIONS

The Hamiltonian of the electronic subsystem of a cuprate compound can be written in the nodal representation as

$$\hat{H} = \sum_{n, n', \sigma} \varepsilon_0(\mathbf{n} - \mathbf{n}')\hat{c}^\dagger_{n\sigma}\hat{c}_{n'\sigma} + U(0)\sum_n \hat{n}_{n\uparrow}\hat{n}_{n\downarrow} + \frac{1}{2}\sum_{n \neq n'} U(\mathbf{n} - \mathbf{n}')\hat{n}_n\hat{n}_{n'}, \quad (3)$$

where the three-dimensional vector index $\mathbf{n} = (n_1, n_2, n_3)$ numbers crystal lattice nodes; $\hat{c}^\dagger_{n\sigma}$ ($\hat{c}_{n\sigma}$) is the operator of creation (annihilation) of particles (a particle is usually understood in cuprate compounds as a hole with spin σ = ↑, ↓ on node **n**); $\hat{n}_{n\sigma} = \hat{c}^\dagger_{n\sigma}\hat{c}_{n\sigma}$ is the operator of the occupation number of the state with spin σ on node n; $\hat{n}_n = \sum_\sigma \hat{n}_{n\sigma}$; and the Hamiltonian parameters $\varepsilon_0(n - n')$, $U(0)$, and $U(\mathbf{n} - \mathbf{n}')$, which determine the seed dispersion law, the intracenter correlation energy, and the interaction energy between particles on nodes **n** and **n**', respectively, do not depend on the normalization volume. As cuprates are compounds with a layered crystal structure characterized by weak interactions between neighboring conducting copper–oxygen planes, these interactions can be ignored and Hamiltonian (3) can be approximated by the sum of the 2*D* Hamiltonians of the electronic subsystems in copper–oxygen planes. Let us consider one of these planes and determine the operators of creation and annihilation of a particle whose 2*D* momentum is **k** and spin σ with the use of the two-dimensional Fourier transform,

$$\hat{c}^\dagger_{k\sigma} = \frac{1}{\sqrt{N_0}}\sum_n \hat{c}^\dagger_{n\sigma}e^{+i\mathbf{k}\cdot\mathbf{n}}, \qquad \hat{c}_{k\sigma} = \frac{1}{\sqrt{N_0}}\sum_n \hat{c}_{n\sigma}e^{-i\mathbf{k}\cdot\mathbf{n}}, \quad (4)$$

where N_0 is the number of elementary cells in the plane under consideration with the normalization area S and **n** = (n_1, n_2) is the 2*D* vector index that determines the position of the node on the plane and in the exponents; it has the meaning and dimension of the radius vector of the node. The Hamiltonian in the 2*D* momentum representation takes the usual form

$$\hat{H}^{(2D)} = \sum_{k\sigma}[\varepsilon_0(\mathbf{k}) - \mu]\hat{c}^\dagger_{k\sigma}\hat{c}_{k\sigma} + \frac{1}{2}\sum_{kk'q\sigma\sigma'} V(\mathbf{q})\hat{c}^\dagger_{k+q,\sigma}\hat{c}^\dagger_{k'-q,\sigma'}\hat{c}_{k'\sigma'}\hat{c}_{k\sigma}. \quad (5)$$

This Hamiltonian is obtained from the $\hat{H} - \mu\hat{N}$ operator, where $\hat{N}$ is the operator of the total number of particles,

$$\varepsilon_0(\mathbf{k}) = \sum_n \varepsilon_0(n)e^{-i\mathbf{k}\cdot\mathbf{n}} \quad (6)$$

can be treated as the seed dispersion law, and the matrix element of interparticle interaction is

$$V(\mathbf{k}) = \frac{1}{N_0}\sum_n U(\mathbf{n})e^{-i\mathbf{k}\cdot\mathbf{n}} \longrightarrow \frac{1}{S}\int U(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}}d^2r. \quad (7)$$

Here, the integral representation corresponds to the case when the $U(\mathbf{n})$ parameters comparatively weakly vary at distances on the order of the lattice constant a. The use of the integral representation of $V(\mathbf{k})$ matrix elements virtually excludes extremely strong intracenter correlations from consideration. As has been

mentioned, these correlations are moderate (comparable with the energy bandwidth t, $U(0) \sim t$) in cuprates.

The existence of a "large" Fourier contour determined by the total concentration of particles $(1 - x)$, where x is the level of doping, follows from angular resolution photoemission spectroscopy measurements [19, 20]. We can therefore treat the state that corresponds to the fully filled one-particle states inside the Fermi contour and fully vacant states outside it as the ground state $|F\rangle$ of Hamiltonian (5). The excited state corresponding to a particle with the momentum $\mathbf{k}$ and spin α added to a system of N particles can then be defined as $\hat{c}^{\dagger}_{k,\alpha}|F\rangle$, and the excited state corresponding to a particle with the momentum $\mathbf{k}$ and spin α removed from the system of N particles, as $\hat{c}_{k,\alpha}|F\rangle$. These excited states are not eigenstates of Hamiltonian (5), because the commutators between this Hamiltonian and the $\hat{c}^{\dagger}_{k,\alpha}$ and $\hat{c}_{k,\alpha}$ operators contain products of three Fermi operators. The linearization of these commutators in the approximation of chaotic phases yields the equations of motion [21]

$$\begin{aligned}[\hat{H}^{(2D)}, \hat{c}^{\dagger}_{k,\alpha}]_{-} &= +(\varepsilon(\mathbf{k}) - \mu)\hat{c}^{\dagger}_{k,\alpha},\\ [\hat{H}^{(2D)}, \hat{c}_{k,\alpha}]_{-} &= -(\varepsilon(\mathbf{k}) - \mu)\hat{c}_{k,\alpha},\end{aligned} \tag{8}$$

which determine the dispersion law

$$\varepsilon(\mathbf{k}) = \varepsilon_0(\mathbf{k}) + \sum_{pq\sigma} V(\mathbf{q})\bar{n}_{\sigma}(p)[\delta q_0 - \delta_{p,k+q}\delta_{\sigma\alpha}], \tag{9}$$

reproducible in principle from angular resolution photoelectron spectra. Here,

$$\bar{n}_{\sigma}(\mathbf{p}) = \langle F|\hat{c}^{\dagger}_{p,\sigma}\hat{c}_{p,\sigma}|F\rangle$$

is the mean occupation number of the state with momentum p and spin σ, which, in the absence of a magnetic field, is independent of the spin index.

The first equation in (8) determines a quasi-particle with the energy $(\varepsilon(\mathbf{k}) - \mu)$ and momentum $k > k_F$, where k_F is the Fermi momentum in the direction of momentum $\mathbf{k}$. The second equation determines a quasi-hole with the energy $(\mu - \varepsilon(\mathbf{k})$ and momentum $k < k_F$. The boundary that separates the regions of filled and vacant states in the $2D$ momentum space, that is, the Fermi contour, is determined by the equation $\varepsilon(\mathbf{k}) = \mu$.

Next, let us consider singlet excitations in the form of a pair of particles outside, or a pair of holes inside, the Fermi contour. The total momentum of the pair $\mathbf{K} = \mathbf{k}_+ + \mathbf{k}_-$ remains unchanged, and the state of the pair at given $\mathbf{K}$ is formed as a linear combination of relative motion states with momenta $\mathbf{k} = (\mathbf{k}_+ - \mathbf{k}_-)/2$,

$$\begin{aligned}|N+2; \mathbf{K}\rangle &= \sum_{k\alpha}\psi^{(+)}_K(\mathbf{k})\hat{c}^{\dagger}_{\frac{K}{2}+k,\alpha}\hat{c}^{\dagger}_{\frac{K}{2}-k,\bar{\alpha}}|F\rangle,\\ |N-2; \mathbf{K}\rangle &= \sum_{k\alpha}\psi^{(-)}_K(\mathbf{k})\hat{c}_{\frac{K}{2}+k,\alpha}\hat{c}_{\frac{K}{2}-k,\bar{\alpha}}|F\rangle,\end{aligned} \tag{10}$$

where the spin index $\bar{\alpha}$ is opposite to α; that is, if $\alpha = \uparrow(\downarrow)$, then $\bar{\alpha} = \downarrow(\uparrow)$. Calculations of the commutators between Hamiltonian (5) and the operators $\hat{A}^{\dagger}_K$ and $\hat{A}_K$ defined as

$$|N+2; \mathbf{K}\rangle = \hat{A}^{\dagger}_K|F\rangle, \quad |N-2; \mathbf{K}\rangle = \hat{A}_K|F\rangle, \tag{11}$$

and subsequent linearization of the resulting equations in the approximation of chaotic phases lead to integral equations that determine the $\psi^{(+)}_K(\mathbf{k})$ and $\psi^{(-)}_K(\mathbf{k})$ coefficients in linear combinations (10). The $\psi^{(+)}_K(\mathbf{k})$ coefficient is given by

$$\begin{aligned}&2\xi_K(\mathbf{k})\psi^{(+)}_K(\mathbf{k}) + \sum_{k'} V(\mathbf{k}-\mathbf{k}')[1 - \bar{n}_{\alpha}(\mathbf{K}/2 + \mathbf{k}')\\ &- \bar{n}_{\bar{\alpha}}(\mathbf{K}/2 - \mathbf{k}')]\psi^{(+)}_{\mathbf{K}}(\mathbf{k}') = (E - 2\mu)\psi^{(+)}_K(\mathbf{k}),\end{aligned} \tag{12}$$

where the kinetic energy of a pair of particles counted from the chemical potential is

$$2\xi_K(\mathbf{k}) = \varepsilon(\mathbf{K}/2 + \mathbf{k}) + \varepsilon(\mathbf{K}/2 - \mathbf{k}) - 2\mu \tag{13}$$

and E is the pair excitation energy to be determined. The energy of particles $\varepsilon(\mathbf{K}/2 \pm \mathbf{k})$ is given by (9). At zero temperature, the occupation numbers in (12) are one or zero if their arguments $\mathbf{K}/2 \pm \mathbf{k}'$ are inside or outside the Fermi contour, respectively. The expression in square brackets therefore equals one if the $\mathbf{K}/2 \pm \mathbf{k}$ momenta of particles that constitute a pair with the total momentum $\mathbf{K}$ are both outside the Fermi contour. This condition determines the momentum space region $\Xi^{(+)}_K$ where the summation over k' in (12) is actually performed. At a nonzero temperature, the multiplier in the square brackets is smaller than one and the summation (with the corresponding weight) is over the whole momentum space. It follows that the interaction between pair components experiences "temperature weakening."

The equation for $\psi^{(-)}_K(\mathbf{k})$ is obtained in the form

$$\begin{aligned}&-2\xi_K(\mathbf{k})\psi^{(-)}_K(\mathbf{k}) + \sum_{k'} V(\mathbf{k}-\mathbf{k}')[1 - \bar{p}_{\bar{\alpha}}(\mathbf{K}/2 + \mathbf{k}')\\ &- \bar{p}_{\alpha}(\mathbf{K}/2 - \mathbf{k}')]\psi^{(-)}_K(\mathbf{k}') = (2\mu - E)\psi^{(-)}_K(\mathbf{k}),\end{aligned} \tag{14}$$

where $\xi_K(\mathbf{k}) < 0$ and the mean occupation numbers of hole states are given by

$$\bar{p}_\alpha(\mathbf{k}) = 1 - \bar{n}_{\bar{\alpha}}(\mathbf{k}). \quad (15)$$

At zero temperature, the expression in square brackets in (14) equals one in the momentum space region $\Xi_K^{(-)}$ inside the Fourier contour, for which the summation is actually performed over the momenta of relative pair motion.

The kinetic energy (13) of a pair with the total momentum $\mathbf{K}$ is the sum of the kinetic energy of relative motion (1) and the energy of the center of mass of the pair that remains unchanged. As it is the relative motion of the pair that is responsible for bound state formation, we can, by defining the relative motion energy as $E_K^{(\lambda)} = \lambda[E - 2\varepsilon(\mathbf{K}/2)]$, reduce (12) and (14) to

$$[E_K^{(\lambda)} - \varepsilon_K^{(\lambda)}(\mathbf{k})]\psi_K^{(\lambda)}(\mathbf{k}) = \sum_{k'} V(\mathbf{k}-\mathbf{k}')\Theta_K^{(\lambda)}(\mathbf{k}')\psi_K^{(\lambda)}(\mathbf{k}'), \quad (16)$$

where $\lambda = \pm 1$ for a pair of particles and holes, respectively, the $\Theta_K^{(\lambda)}(\mathbf{k})$ characteristic function has the form

$$\Theta_K^{(+)}(\mathbf{k}) = 1 - \bar{n}_\alpha(\mathbf{K}/2 + \mathbf{k}) - \bar{n}_{\bar{\alpha}}(\mathbf{K}/2 - \mathbf{k}),$$
$$\Theta_K^{(-)}(\mathbf{k}) = 1 - \bar{p}_{\bar{\alpha}}(\mathbf{K}/2 + \mathbf{k}) - \bar{p}_\alpha(\mathbf{K}/2 - \mathbf{k}), \quad (17)$$

and the kinetic energy of a pair of particles or holes is $\varepsilon_K^{(\lambda)}(\mathbf{k}) = \lambda\varepsilon_K^{(r)}(\mathbf{k})$.

4. THE RELATIVE MOTION OF A PAIR

Let us rewrite (16) in the form of the integral equation

$$\psi(\mathbf{k}) = G(\mathbf{k}; E)\int U(\mathbf{k}-\mathbf{k}')\psi(\mathbf{k}')\Theta(\mathbf{k}')d^2k', \quad (18)$$

where $\psi(\mathbf{k}) \equiv \psi_K^{(\lambda)}(\mathbf{k})$, $U(\mathbf{k} - \mathbf{k}') \equiv S/(2\pi)^2 V(\mathbf{k} - \mathbf{k}')$, $\Theta(\mathbf{k}') \equiv \Theta_K^{(\lambda)}(\mathbf{k}')$, the integration is over the whole $2D$ Brillouin zone, and the retarded Green function of the free relative motion of a pair is given by

$$G(\mathbf{k}) = \frac{1}{E - \varepsilon_K^{(\lambda)}(\mathbf{k}) + i\gamma\,\mathrm{sgn}\,E}, \quad (19)$$

where $\gamma \longrightarrow +0$.

Consider an auxiliary homogeneous linear Fredholm equation of the second kind with the symmetrical kernel $U(\mathbf{k} - \mathbf{k}')$, which determines the eigenfunctions $\varphi_s(\mathbf{k})$ and eigenvalues λ_s of this kernel,

$$\varphi_s(\mathbf{k}) = \lambda_s\int U(\mathbf{k}-\mathbf{k}')\varphi_s(\mathbf{k}')d^2k'. \quad (20)$$

If the kernel is nondegenerate, the eigenfunctions $\varphi_s(\mathbf{k})$ form an infinite complete orthonormal system of functions [22],

$$\int\varphi_{s'}^*(\mathbf{k})\varphi_s(\mathbf{k})d^2k = \delta_{s's}, \quad (21)$$

and the kernel of the integral operator in (20) can be represented as the Hilbert–Schmidt spectral decomposition in these functions [22]

$$U(\mathbf{k}-\mathbf{k}') = \sum_s \frac{\varphi_s(\mathbf{k})\varphi_s^*(\mathbf{k}')}{\lambda_s}. \quad (22)$$

Let us expand the $\Theta(\mathbf{k})\psi(\mathbf{k})$ function in terms of the complete system of functions $\varphi_s(\mathbf{k})$,

$$\Theta(\mathbf{k})\psi(\mathbf{k}) = \sum_s \alpha_s\varphi_s(\mathbf{k}). \quad (23)$$

The expansion coefficients are written as

$$\alpha_s = \int\varphi_s^*(\mathbf{k})\psi(\mathbf{k})\Theta(\mathbf{k})d^2k. \quad (24)$$

We can use (22) and (24) to reduce integral equation (18) to an infinite system of linear homogeneous equations

$$\sum_{s'}\left\{\delta_{ss'} - \frac{1}{\lambda_{s'}}G_{ss'}(E)\right\}\alpha_{s'} = 0, \quad (25)$$

in which the retarded Green function in the representation formed by the $\varphi_s(\mathbf{k})$ functions is written as

$$G_{ss'}(E) = \int\varphi_s^*(\mathbf{k})G(\mathbf{k}; E)\Theta(\mathbf{k})\varphi_{s'}(\mathbf{k})d^2k. \quad (26)$$

Note that the $\Theta(\mathbf{k})$ function at zero temperature equals one inside some momentum space region Ξ ($\Xi = \Xi_K^{(+)}$ for a pair of particles and $\Xi = \Xi_K^{(-)}$ for a pair of holes) and zero outside this region. For this reason, the integration in momentum is then actually performed over this bounded momentum space region. The condition

of nontrivial compatibility of equations (25) determines the energy spectrum for the relative motion of a pair.

If the kernel is degenerate, algebraic system (25) reduces to a finite system of linear homogeneous equations, and the condition of its compatibility takes the form

$$\det\left\{\delta_{ss'} - \frac{1}{\lambda_{s'}}G_{ss'}(E)\right\} = 0, \tag{27}$$

where index s numbers a finite set of eigenfunctions of the degenerate kernel.

It follows from definition (26) that the main contribution to the matrix elements of the Green function is made by a comparatively small momentum space region Ξ. This gives grounds for replacing the true nondegenerate kernel $U(\mathbf{k} - \mathbf{k}')$ by some degenerate kernel $U_d(\mathbf{k} - \mathbf{k}')$, which is close to the true one within region Ξ. Let k_b be the characteristic linear size of region Ξ, such that the main contribution to (18) is made by momenta $k \le k_b$ and $k' \le k_b$. The vector of momentum transfer in scattering $\boldsymbol{\kappa} = \mathbf{k} - \mathbf{k}'$ is then limited, $\kappa \le 2k_b$, and the $U(\kappa)$ kernel in the region $0 \le \kappa \le 2k_b$ can be approximated by the simple quadratic function

$$U_d(\kappa) = U_0 r_0^2(1 - \kappa^2 r_0^2/2), \tag{28}$$

where $U_0 > 0$ and r_0 are the parameters of the repulsive interaction potential between particles that characterize its amplitude and range of action, respectively. At $\kappa > 2k_b$, we must set $U_d(\kappa) = 0$. Treating this approximate potential as the kernel of integral equation (20), we can find the corresponding eigenfunctions and eigenvalues. A finite number of eigenfunctions correspond to degenerate kernel (28), and it is easy to see that this kernel has four eigenfunctions only.

The eigenfunctions of a degenerate kernel reproduce its form [22] and can be written as

$$\varphi(\mathbf{k}) = a + \boldsymbol{\chi}\mathbf{k} + bk^2 \tag{29}$$

if $0 \le k \le k_b$ and as $\varphi(\mathbf{k}) = 0$ if $k > k_b$, where the coefficients a, b, and $\boldsymbol{\chi}$ must be determined for each of the eigenfunctions. The equations that determine these coefficients are considerably simplified because the region of integration Ξ in (20) is invariant with respect to inversion, $\mathbf{k} \longrightarrow -\mathbf{k}$. It follows that the eigenfunctions possess certain parity with respect to variable $\mathbf{k}$. The system of equations that determine the coefficients in (29) is divided into two independent subsystems, one of which gives even, and the other odd, eigenfunctions.

Let the k_1 and k_2 coordinate axes in the momentum space be directed along the symmetry axes of region Ξ. We will use the notation

$$K_n = r_0^{2n+2}\int_\Xi k^{2n} d^2k, \qquad K_n^{(s)} = r_0^{2n+2}\int_\Xi k_s^{2n} d^2k, \tag{30}$$

where $n = 0$, 1, and 2 and $s = 1$, 2 corresponds to a Cartesian coordinate axis in the momentum space. The normalized odd eigenfunctions and the corresponding eigenvalues of the degenerate nucleus are immediately found as

$$\varphi_s^{(-)}(\mathbf{k}) = \frac{r_0^2 k_s}{\sqrt{K_1^{(s)}}}, \quad \lambda_s^{(-)} = \frac{1}{U_0 K_1^{(s)}}. \tag{31}$$

The condition of the solvability of the system of equations with respect to the a and b coefficients that determine even eigenfunctions gives two eigenvalues,

$$\lambda_s^{(+)} = \frac{-2}{U_0 K^2}[(K_0 - K_1) \pm \sqrt{(K_0 - K_1)^2 + K^2}] \tag{32}$$

and the corresponding real eigenfunctions,

$$\varphi_s^{(+)}(\mathbf{k}) = a_s\left[1 - \frac{K_0 \pm \sqrt{(K_0 - K_1)^2 + K^2}}{2K_1 - K_2} r_0^2 k^2\right], \tag{33}$$

where, as previously, $s = 1$, 2, a_s are the normalization coefficients, and, by definition,

$$K^2 = K_0 K_2 - K_1^2 \tag{34}$$

is a positive value (by virtue of the Cauchy–Bunyakowsky inequality for the K_n integrals). Note that the eigenfunctions of degenerate kernel (28) that we found are orthogonal with the weight $\Theta(\mathbf{k})$.

Also note that one of the eigenvalues (32) that corresponds to the upper sign of the root is negative, whereas the other one and all eigenvalues (31) are positive. The sign of the eigenfunction (33) that corresponds to the negative eigenvalue is not constant in the region of its definition and alternates on a certain line (an arc of a circle) inside this region.

Not any kernel that corresponds to a repulsive (positive definite) potential can have negative eigenvalues (or at least one negative eigenvalue). For instance, all eigenvalues of the kernel described by the Gauss function are nonnegative.

It follows that the 4×4 matrix of Green operator (26) splits into two 2×2 matrices between even $G_{ss'}^{(+)}(E)$ and odd $G_{ss'}^{(-)}(E)$ eigenfunctions of degenerate kernel (28). The latter matrix is diagonal, and its elements are

$$G_s^{(-)}(E) = \int G(\mathbf{k}; E)[\varphi_s^{(-)}(\mathbf{k})]^2 \Theta(\mathbf{k}) d^2k. \quad (35)$$

Equation (27) therefore splits into three independent equations,

$$\det\{\lambda_{s'}^{(+)}\delta_{ss'} - G_{ss'}^{(+)}(E)\} = 0, \quad G_s^{(-)}(E) = \lambda_s^{(-)}, \quad (36)$$

where s, $s' = 1, 2$, and

$$G_{ss'}^{(+)}(E) = \int G(\mathbf{k}; E)\varphi_s^{(+)}(\mathbf{k})\varphi_{s'}^{(+)}(\mathbf{k})\Theta(\mathbf{k}) d^2k. \quad (37)$$

These equations determine the energy spectrum of the relative motion of a pair of particles.

5. STATIONARY STATES

The reference point for counting energy can conveniently be changed in calculations as $E \longrightarrow E + \varepsilon_p$ to pass from the energy of relative motion to the excitation energy by putting $\varepsilon_K^{(\lambda)}(\mathbf{k}) = \zeta(\mathbf{k}) + \varepsilon_p$ in (19). By definition, the excitation energy of a pair belongs to the continuous spectrum band, $0 \le \zeta(\mathbf{k}) \le \zeta_0$. When the reference point is selected this way, the bound state energy is negative, $E < 0$. Accordingly, the energy denominator in (19) has a definite sign, $E - \zeta(\mathbf{k}) < 0$, and the infinitesimal imaginary term in the denominator of the Green function in the integrand in (19) can be ignored. It follows that we have $G_s^{(\pm)} < 0$ if $E < 0$. As both eigenvalues corresponding to odd eigenfunctions are positive at $U_0 > 0$ [see (31)], the conclusion can be drawn that the equations $G_s^{(-)}(E) = \lambda_s^{(-)}$ do not have solutions at $E < 0$.

The equation from (36) that gives even eigenfunctions can be written in the form

$$G_{11}^{(+)}G_{22}^{(+)} - [G_{12}^{(+)}]^2 = \lambda_2^{(+)}G_{11}^{(+)} + \lambda_1^{(+)}G_{22}^{(+)} - \lambda_1^{(+)}\lambda_2^{(+)}. \quad (38)$$

A solution to (38) can conveniently be found graphically if its left- and right-hand sides are treated as functions of the variable $E < 0$. The function on the left-hand side of (38) is nonnegative by virtue of the Cauchy–Bunyakowsky inequality for the corresponding integrals. The sum of the first two terms on the right-hand side of (38) is also nonnegative in spite of the different signs of $\lambda_1^{(+)}$ and $\lambda_2^{(+)}$. Indeed, we can put $\mathbf{k} = \mathbf{k}'$ in (22) and calculate the sum on the right-hand side of (22) for degenerate kernel (28); this sum then consists of only four terms. Using the explicit form of the odd eigenfunctions and the corresponding eigenvalues (31), we find

$$[\varphi_1^{(+)}(\mathbf{k})]^2/\lambda_1^{(+)} + [\varphi_2^{(+)}(\mathbf{k})]^2/\lambda_2^{(+)} = U_d(k\sqrt{2}). \quad (39)$$

The right-hand side of (39) is essentially positive over the whole region where the relative pair motion momentum is defined because of the condition $U_d(|\mathbf{k} - \mathbf{k}'|) > 0$, which corresponds to repulsion between particles. Equation (39) can be used to represent the first two terms on the right-hand side of (38) as the product of the negative value

$$\lambda_1^{(+)}\lambda_2^{(+)} = -4/U_0^2K^2 < 0 \quad (40)$$

and the integral of form (37), in which we have the obvious inequality $G(\mathbf{k}; E) < 0$ if $E < 0$ and the product $\varphi_s^{(+)}(\mathbf{k})\varphi_{s'}^{(+)}(\mathbf{k})$ is replaced by the definitely positive $U_d(k\sqrt{2})$ value,

$$\lambda_2^{(+)}G_{11}^{(+)} + \lambda_1^{(+)}G_{22}^{(+)} = \int \frac{4U_d(k\sqrt{2})\Theta(\mathbf{k})}{U_0^2K^2[|E| + \zeta(\mathbf{k})]} d^2k. \quad (41)$$

Note that the value (41) is inversely proportional to the effective coupling constant U_0.

The asymptotic behavior of Green functions (37) as $E \longrightarrow -\infty$ follows from their definition,

$$G_{ss'}^{(+)}(E) \propto E^{-1}\delta_{ss'}. \quad (42)$$

The solution to (38) that corresponds to extremely large effective coupling constant U_0 values then takes the form

$$E = -\frac{U_0}{2}[\sqrt{(K_0 - K_1)^2 + K^2} - (K_0 - K_1)]. \quad (43)$$

Note that the left-hand side of (38) tends to zero as E^{-2} when $E \longrightarrow -\infty$, whereas the right-hand side tends to a positive finite limit $|\lambda_1^{(+)}\lambda_2^{(+)}|$.

Green functions (37) are given by integrals over the $\Xi_K^{(-)}$ and $\Xi_K^{(+)}$ regions. To consider their asymptotic behavior as $E \longrightarrow -0$, we can conveniently pass from the integration in the components of the momentum of relative motion to the integration in new variables ζ and τ, where ζ is the excitation energy and τ is the curvilinear coordinate in the space of momenta along the ζ = const isoline of the energy of relative pair motion. Functions (37) can then be rewritten as

$$G_{ss'}^{(+)}(E) = \int_0^{\zeta_0} G(\zeta; E)g(\zeta)F_{ss'}(\zeta)d\zeta, \quad (44)$$

where $g(\zeta)$ is the density of states of relative pair motion corresponding to the $\Xi_K^{(-)}$ or $\Xi_K^{(+)}$ region (fur-

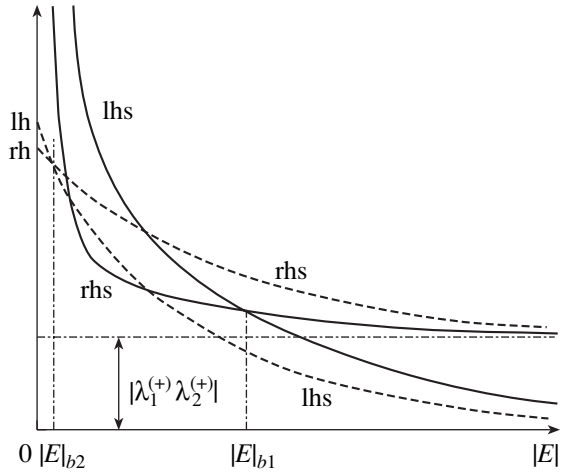

Fig. 4. Character of solutions to (38) at a finite pair Fermi contour length (solid lines) and when the pair Fermi contour degenerates into points (dashed lines). The curves describing the left- and right-hand sides of (38) are denoted by lhs and rhs, respectively. Solutions at finite ($|E|_{b1}$) and zero ($|E|_{b2}$) pair Fermi contour lengths are shown. The solution for zero pair Fermi contour length corresponds to the condition $U_0 > U_0^{(c)}$. Otherwise, the value of the left-hand side of (38) (denoted by lh on the axis of ordinates) becomes smaller than the value of the right-hand side (rh), and (38) has no solutions.

ther denoted by Ξ), ζ_0 is the width of the continuous spectrum band with respect to motion, and

$$F_{ss'}(\zeta) = \frac{1}{g(\zeta)} \int\limits_{\zeta = \text{const}} \frac{\varphi_s^{(+)}(\mathbf{k})\varphi_{s'}^{(+)}(\mathbf{k})}{|\nabla_k \zeta(\mathbf{k})|}\Theta(\mathbf{k})d\tau. \quad (45)$$

It follows that $F_{ss'}(\zeta)$ is the mean values of the $\varphi_s^{(+)}(\mathbf{k})\varphi_{s'}^{(+)}(\mathbf{k})$ products on the ζ = const isoline. Note that the $F_{ss}(\zeta)$ functions at $s = s'$ are necessarily positive, like their combination

$$F^2 \equiv F_{11}F_{22} - F_{12}^2$$

is (by virtue of the Cauchy–Bunyakowsky inequality). Here and throughout, $F_{ss'} \equiv F_{ss'}(0)$.

Green functions (37) have a singularity at small $|E|$, which can be examined by the well-known Kantorovich method [23]. Suppose that the function

$$f(\zeta) = g(\zeta)F_{ss'}(\zeta)$$

is differentiable at $E = 0$. Equation (44) can then be written in the form

$$G_{ss'}^{(+)}(E) = -f(0)\ln\frac{\zeta_0}{|E|} + f'(0)|E|\ln\frac{\zeta_0}{|E|} - \left\{ f'(0)\zeta_0 + \int\limits_0^{\zeta_0} \frac{[f(\zeta) - f(0) - f'(0)\zeta]}{\zeta + |E|}d\zeta \right\}, \quad (46)$$

where primes denote the differentiation with respect to ζ. The integral in braces does not have a singularity at $|E| = 0$. At small $|E|$ values, the whole expression in braces can therefore be replaced by its value at $|E| = 0$; we denote this value by $A_{ss'}$. It follows that $-A_{ss'}$ is the periodic part of the Green function when $|E| \longrightarrow 0$. Therefore, the Green functions at $|E| \ll \zeta_0$ can be represented as

$$G_{ss'}^{(+)}(E) = -[g(0)F_{ss'}(0)]\ln\frac{\zeta_0}{|E|} + [g(0)F_{ss'}(0)]'|E|\ln\frac{\zeta_0}{|E|} - A_{ss'}. \quad (47)$$

The first two terms in (47) determine the character of the singularity of the Green functions at point $E = -0$. If the density of states on the pair Fermi contour is nonzero (the pair Fermi contour has a finite length), that is, if $g(0) \neq 0$, the Green functions exhibit a logarithmic singularity, then $G_{ss'}^{(+)}(E) \propto \ln|E|$ as $E \longrightarrow -0$. However, if the pair Fermi contour degenerates into points and $g(0)$ vanishes, the character of the Green function singularity as $E \longrightarrow -0$ becomes much softer, $G_{ss'}^{(+)}(E) \propto |E|\ln 1/|E|$.

If the pair Fermi contour has a finite length, the left-hand side of (38) diverges as $[\ln(1/|E|)]^2$ when $E \longrightarrow -0$, whereas the right-hand side only diverges as $[\ln(1/|E|)]$. Taking into account the asymptotic behavior of the Green functions as $E \longrightarrow -\infty$ considered above, we obtain the functions representing the left- and right-hand sides of (38) in the form schematically shown in Fig. 4 by solid lines. According to this figure, (38) has a unique solution no matter what the effective coupling constant U_0 value. The character of this solution in the limit of weak coupling is easy to determine using the explicit equations for Green functions (47). When considering the case $g(0) \neq 0$, the second term, which vanishes at $|E| = 0$, should be omitted. Equation (38) transforms into a quadratic equation with respect to $x \equiv g(0)\ln(\zeta_0/|E|)$, whose positive solution can be represented as $x_0 = 1/cU_0$, where c is a positive constant, which, in the weak coupling limit, is given by

$$c = C\left[\sqrt{1 + \frac{F^2K^2}{C^2}} + 1\right]. \quad (48)$$

Here, the C constant, which is independent of the effective coupling constant U_0, is determined by the integral along the pair Fermi contour

$$C = \frac{1}{g(0)U_0} \int\limits_{\zeta = 0} \frac{U_d(k\sqrt{2})\Theta(\mathbf{k})}{|\nabla_k \zeta(\mathbf{k})|}d\tau. \quad (49)$$

We therefore have

$$|E| = \zeta_0 \exp\left(-\frac{2}{cU_0 g(0)}\right), \quad (50)$$

that is, the bound state of the relative motion of a pair is already formed at an arbitrarily small effective coupling constant value. Note that, if $FK \ll C$, (50) retains its form with the replacement $c \longrightarrow C$.

If the pair Fermi contour degenerates into points (the case considered in [5]), the first term in (47) disappears and the contribution of the periodic part to the Green functions becomes predominant as $|E| \longrightarrow 0$. A solution to (38) then only exists for a fairly high effective coupling constant U_0 value that exceeds some $U_0^{(c)}$ value (see Fig. 4). Indeed, if $U_0 < U_0^{(c)}$, the left-hand side of (38) denoted by lh in Fig. 4 becomes smaller than the right-hand side (rh) at $|E| = 0$; the curves corresponding to the left- (lhs) and right-hand (rhs) sides of (38) [these curves are shown in Fig. 4 by dashed lines for $U_0 > U_0^{(c)}$ then do not intersect, and there is therefore no solution to (38).

Solution (50) found for repulsive interaction of form (28) for the pair Fourier contour of a finite length [$g(0) \neq 0$] formally coincides with the solution to the Cooper problem for attractive interaction of the form $U(\mathbf{k} - \mathbf{k}') = -U_0$ (U_0 is a positive constant, and interaction is nonzero in an energy layer ζ_0 wide) of a pair of particles with zero total momentum [1]. This solution exists for pairs of particles excited in the $\Xi^{(+)}$ region and pairs of holes excited in the $\Xi^{(-)}$ region. According to [24], the bound state energy of a Cooper pair has exponential smallness (factor 2 in the exponent) compared with the superconducting gap width. Of course, the same refers to pairing with a large total momentum (factor 2 in the exponent in (50) compared with the order parameter calculated in [13]). The reason for this smallness is as follows. By virtue of the problem statement itself, the solution to (50) is obtained against the background of the nonrearranged ground state of the normal Fermi liquid without taking into account scattering between momentum space regions $\Xi^{(+)}$ and $\Xi^{(-)}$. The very existence of bound state (50) is evidence of instability of the ground state of the electronic system treated as the state that corresponds to filling all states inside the momentum space region bounded by the Fermi contour [25].

Note that bound stationary state (50) appears in repulsive interactions if the kernel of integral equation (18) has at least one negative eigenvalue. As has been shown above, kernel (29) certainly has such a negative eigenvalue. The passage to the $r_0^2 \longrightarrow 0$ limit in (29) at $U_0 r_0^2 = \text{const}$ causes the appearance of a degenerate kernel with a single positive eigenvalue. In such a passage, the pole of the amplitude of scattering in the complex energy plane corresponding to the bound stationary state shifts from the real axis to the upper half-plane and, at a hyperbolic momentum space metric, formally corresponds to some state characterized by "negative decay" [11]. This can also be considered evidence of instability of the ground state in the form of Fermi filling.

The passage to the $r_0^2 \longrightarrow 0$, $U_0 r_0^2 = \text{const}$ limit, which formally corresponds to the zero-radius potential approximation [26], admits a simple phenomenology, which allows the pair bound state energy obtained by solving (18) with a potential of form (28) to be fixed if the "memory" of this potential is retained in the discontinuous piecewise-constant wave function of the pair, which abruptly changes sign on some line (line of zeros) inside the region of relative motion momentum definition. Such a solution for the superconducting order parameter that appears for repulsive interaction and pairing with a large total momentum was obtained in [11].

It is well known [6] that a purely imaginary vertex function pole in the upper complex energy half-plane corresponds to the bound state in the Bardin–Cooper–Schriffer model. The absolute value of this pole gives the energy gap in the spectrum of the excited system that appears as a result of the rearrangement of its ground state. An examination of the vertex function for the class of repulsive potentials (28) with one or several negative eigenvalues appears to give a similar result. This assertion can be proved the same way as for the repulsive potential with a negative scattering length for one of the angular momentum values $l \neq 0$ [27, Section 54], which therefore belongs to the class of the potentials under consideration. Provided that the equal sums of entering and exiting momenta in the vertex function are the pair momentum $\mathbf{K}$, the purely imaginary vertex function pole gives the energy gap value in the spectrum of excitations of relative pair motion. As in the Bardin–Cooper–Schriffer model, the dependence of the gap found in [13] is described in the weak coupling limit by an exponential function of type (50) without the multiplier 2 in the exponent and with a preexponential factor that is twice as large.

6. QUASI-STATIONARY STATES

At excitation energies $0 < E < \zeta_0$, integral equation (18) determines the states of the continuous spectrum of relative pair motion. The distribution of energy levels in the continuous spectrum band $0 < \zeta < \zeta_0$ can be found from system (36). It may well be that the interaction of particles constituting a pair results in a substantial crowding of relative motion energy levels in a small neighborhood of some point ζ_q, $0 < \zeta_q < \zeta_0$ within the continuous spectrum band. The corresponding change in the density of relative motion states is resonance in character, and the arising peak of the density of states resembles a smeared δ-shaped peak that corresponds to

the bound state. Such a quasi-stationary state is characterized by a finite lifetime τ_q, which is inversely proportional to the extent of smearing of the resonance peak.

The complex solution to (36) that corresponds to the quasi-stationary state has the form $\tilde{E} = E - i\Gamma/2$, where E is the resonance energy and $\Gamma \propto \hbar/\tau_q > 0$ is the quasi-stationary state damping. At a finite damping value, the infinitesimal term in the denominator of Green function (19) can be ignored. Omitting the superscript ($\pm$), Green functions (35) and (37) can then be written as

$$G_{ss'}(E, \Gamma) = G'_{ss'}(E, \Gamma) + iG''_{ss'}(E, \Gamma),$$

where the real and imaginary parts have the form

$$G'_{ss'}(E, \Gamma) = \int_0^{\zeta_0} \frac{(E-\zeta)F_{ss'}(\zeta)g(\zeta)}{(E-\zeta)^2 + \Gamma^2/4} d\zeta, \quad (51)$$

$$G''_{ss'}(E, \Gamma) = \frac{\Gamma}{2}\int_0^{\zeta_0} \frac{F_{ss'}(\zeta)g(\zeta)}{(E-\zeta)^2 + \Gamma^2/4} d\zeta. \quad (52)$$

Here, the $F_{ss'}(\zeta)$ functions are defined by (45). It follows that the diagonal matrix elements (52) are positive definite.

System (36), which now determines two unknown values E and Γ, should be written in the form of an extended system of equations for its real and imaginary parts. We obtain $G'_s = \lambda_s^{(-)}$ and $G''_s = 0$ for all Eqs. (36) related to odd eigenfunctions. It follows that neither discrete relative motion energy levels that fall out of the continuous spectrum nor the quasi-stationary states in the continuous spectrum band can be related to odd eigenfunctions. Indeed, as follows from definition (52), the equation $G''_s = 0$ cannot have solutions at any $\Gamma > 0$.

The real and imaginary parts of (36) related to even eigenfunctions have the form

$$G'_{11}G'_{22} - G'^2_{12} = \lambda_2^{(+)}G'_{11} + \lambda_1^{(+)}G'_{22}$$
$$- \lambda_1^{(+)}\lambda_2^{(+)} + G''_{11}G''_{22} - G''^2_{12}, \quad (53)$$
$$G'_{11}G''_{22} + G''_{11}G''_{22} - 2C'_{12}G''_{12} = \lambda_2^{(+)}G''_{11} + \lambda_1^{(+)}G''_{22}.$$

Every equation in (53) determines some curve on the E, $\Gamma/2$ plane. The intersection points between these curves, if exist, give solutions to (53).

Let us consider the quasi-stationary states mainly formed by one-particle states in the immediate vicinity of the pair Fourier contour. The excitation energy of quasi-stationary states can then be taken to be a small value. In addition, we only consider the case when the quasi-stationary state damping is also small, namely, we assume that $\Gamma \le 2E$. If E and Γ are small, it is easy to examine the singular contribution to the real part of Green functions (51) using the procedure suggested by Kantorovich [23]. We restrict consideration to the pair Fermi contour of a finite length. For the real Green function part [Eq. (51)], we obtain

$$G'_{ss'}(E, \Gamma) \propto -g(0)F_{ss'}(0)\ln\sqrt{\frac{\zeta_0^2}{E^2 + \Gamma^2/4}} - A_{ss'}, \quad (54)$$

where, as in (47), $A_{ss'}$ is the periodic Green function part as $E \longrightarrow +0$. In a similar way, the imaginary part [Eq. (52)], which is a periodic function as $E \longrightarrow +0$, is found in the approximate form

$$G''_{ss'}(E, \Gamma) \propto g(0)F_{ss'}(0)\left[\frac{\pi}{2} + \arctan\frac{2E}{\Gamma}\right] \quad (55)$$

[correcting terms are omitted because the exact limit of (52) as $\Gamma \longrightarrow +0$, which equals $\pi g(E)F_{ss'}(E)$, is close to the corresponding limit of (55) at small E values; the latter equals $\pi g(0)F_{ss'}(0)$].

Functions (54) depend on E and Γ combined as $E^2 + \Gamma^2/4 \equiv \rho^2$. If ρ is treated as a polar radius on the $\Gamma/2$, E plane, then, setting $\Gamma/2 = \rho\cos\varphi$ and $E = \rho\sin\varphi$, where φ is the corresponding polar angle, we obtain $\arctan 2E/\Gamma = \varphi$. The ρ variable can conveniently be replaced by $x = g(0)\ln\zeta_0/\rho$, which is positive definite at $\rho < \zeta_0$.

The second equation in (53) contains no dependence on φ and determines some $x = x_0$ value. The solution to this equation is

$$x_0 = \frac{1}{F^2}\left(\frac{2C}{K^2U_0} - D\right), \quad (56)$$

where $2D \equiv A_{11}F_{22} + A_{22}F_{11} - 2A_{12}F_{12}$ and the positive constant C is given by integral (49). The equality $x = x_0$ determines a circle on the ρ, φ plane; its radius is

$$\rho_0 = \zeta_0\exp\left(-\frac{x_0}{g}\right), \quad (57)$$

where $g \equiv g(0)$. In the weak coupling limit when the second term in parentheses in (56) can be ignored, (57) takes the form

$$\rho_0 = \zeta_0\exp\left(-\frac{2}{c'U_0g}\right), \quad (58)$$

where $c' = C/F^2K^2$.

Excluding x in the first equation in (53), that is, assuming that $x = x_0$ in it, we obtain the equation for the polar angle φ

$$F^2 g^2 \left(\frac{\pi}{2} + \varphi\right)^2 = A^2 + \frac{\mu}{U_0} - \frac{A}{K^2 U_0^2} - F^2 x_0^2. \qquad (59)$$

Here,

$$A^2 = A_{11}A_{22} - A_{12}^2 \geq 0, \quad \mu/U_0 = \lambda_2^{(+)} A_{11} + \lambda_1^{(+)} A_{22}.$$

Equation (59) determines the polar angle $\varphi = \varphi_0$, which should satisfy the inequality $0 \leq \varphi_0 \leq \pi/2$. This angle and the polar radius $\rho = \rho_0$ found above determine the solution to system (53) if it does exist. Note that a key role in the problem of the quasi-stationary state is played by the contributions of the periodic term $A_{ss'}$ to the real parts of Green functions. Ignoring these contributions (as in the problem of bound states) leads to (59) with $A = 0$ and $\mu = 0$. Clearly, (59) then has no solution.

The conditions of the formation of quasi-stationary states with small damping are determined by two inequalities, $x_0 \geq 0$ and $0 \leq \varphi_0 \leq \pi/2$. The first one and the quite obvious condition $D \geq 0$ lead to bounding of the effective coupling constant from above, $U_0 \leq U_0^{(M)} \equiv 2C/DK^2$. If (59) has a solution in the interval $(0, \pi/2)$ under this constraint, this solution $\varphi = \varphi_0$ and (57) determine the quasi-stationary state. Figure 5 shows schematically how a solution corresponding to a quasi-stationary state appears. The vertical B axis corresponds to the values of the left- and right-hand sides of (59), and the horizontal axis, to the effective coupling constant U_0 values. Figure 5 illustrates the character of solving (59) for a pair with some momentum **K**, that is, at definite equation parameter values, which are characteristics of the definition region of the momentum of relative pair motion. As U_0 tends to zero, the right-hand side of (59) tends to $-\infty$ as $-U_0^{-2}$, and, as $U_0 \longrightarrow \infty$, it tends to a constant limit $B_0 = A^2 - D^2/F^2$. The $U_0^{(M)}$ value bounds the effective coupling constant from above. For this reason, while the left-hand side of (59) changes from $B_1 = \pi^2 F^2 g^2/4$ to $B_2 = \pi^2 F^2 g^2$ when the polar angle changes from 0 to $\pi/2$, as is shown on the vertical axis in Fig. 5, the intersection points between the horizontal straight lines $B = B_1$ and $B = B_2$ and the plot of the right-hand side of (59) denoted by rhs determine the interval of the effective coupling constant values

$$U_0^{(m)} \leq U_0 \leq U_0^{(M)}, \qquad (60)$$

within which quasi-stationary states with long lifetimes can be formed.

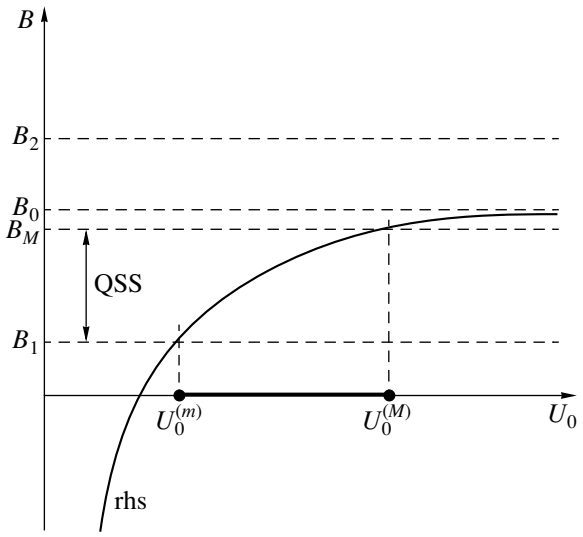

Fig. 5. Graphic illustration of the appearance of a quasi-stationary state depending on the effective coupling constant U_0. The values of the left- and right-hand sides of (59) are plotted on the axis of ordinates (B). The left-hand side values, independent of U_0 and allowed by the condition $0 \leq \varphi_0 \leq \pi/2$, lie within the interval $B_1 \leq B \leq B_2$. The right-hand side is described by the curve labeled rhs. Solutions to (56) and (59) that correspond to quasi-stationary states exist in the interval of effective coupling constant values $U_0^{(m)} \leq U_0 \leq U_0^{(M)}$.

The actual $U_0^{(m)}$ and $U_0^{(M)}$ values substantially depend on the size and shape of the integration region Ξ and on the form of the dispersion law for relative pair motion $\zeta(\mathbf{k})$, which are in turn determined by the electronic dispersion law and the value and direction of the total pair momentum **K**. As asymptotic equations (54) and (55) are valid if the mirror Fourier contour nesting condition is satisfied, the conclusion can be drawn that there exist some set of vectors **K** close to optimal (when the mirror nesting condition is best satisfied), for which the condition of the formation of quasi-stationary states with small damping can certainly be met. Only one of the possibilities of the formation of a quasi-stationary state with some momentum **K** is shown in Fig. 5. A change in **K** entails a change in the parameters of (56) and (59), in particular, it may happen that $B_1 > B_M$, and no quasi-stationary state will be formed. If $B_2 < B_M$, where B_M corresponds to the coupling constant $U_0^{(M)}$ value, the position of the upper boundary of the effective coupling constant is determined by the intersection of the $B = B_2$ horizontal straight line and the plot of the right-hand side of (59).

Note that a change in the effective coupling constant at a given **K** value (for instance, its decrease as the level of doping increases [5]) can displace this constant from

the interval ($U_0^{(m)}$, $U_0^{(M)}$) and thereby cause the disappearance of quasi-stationary states.

7. THE PSEUDOGAP STATE

Currently, there are various points of view on the nature of the pseudogap state based on an analysis of models with strong electron correlation. One of these is related to the suggestion that the pseudogap arises thanks to strong fluctuations of the superconducting phase order parameter [28–30]. This presupposes that noncoherent pair states exist in some temperature interval above the superconducting transition temperature T_C. According to another point of view [31, 32], a special substance phase exists under pseudogap conditions and is characterized by a "latent" (that is, difficult to determine experimentally) order parameter caused by the circulation of orbital currents along chemical bonds in copper–oxygen planes. The staggered order in the distribution of the sign of circulation over unit cells corresponds to antiferromagnetic orbital ordering. A noncoherent state in the form of a fluctuation of such an orbital antiferromagnetic order [33, 34] is also treated as a possible reason for the formation of the pseudogap state of cuprates.

Although the specified approaches to describing the pseudogap state are based on different physical ideas, they eventually lead us to conclude that states in the form of either noncoherent pairs excited from the condensate when a substance passes T_C as its temperature rises (pairing in the particle–particle channel) or pairs formed by orbital currents that circulate in opposite directions in neighboring unit cells [33] (pairing in the particle–hole channel) exist above T_C in a wide temperature interval. The wide temperature range in which phase fluctuations exist at a nonzero amplitude of the superconducting order parameter is explained either by the presence of a quantum critical point [33] or by thermal disordering of the *d* density wave [34] corresponding to antiferromagnetism. Note that, in terms of the *t*–*J* model studied in [33], two Bose branches of the spectrum of elementary excitations are naturally formed. One of these is related to the absence of fermionic excitations, and the other, to filling a node with two fermions. It was shown in [33] that bosons of both kinds (interpreted as holes) condense at the bottom of their energy bands. Zero momentum ($\mathbf{Q}_1 = 0$) corresponds to the position of the bottom of one band and a large momentum [$\mathbf{Q}_2 = (\pi, \pi)$], to the bottom of the second band. The character of correlation of currents that circulate in unit cells shows that holes of two kinds effectively attract each other, which corresponds to a comparatively low binding energy between a pair of holes and various orbital current circulation signs.

Quasi-stationary states with a comparatively long lifetime obtained in terms of the concept of delocalized electrons also correspond to special elementary excitations of the crystal electronic subsystem formed by pairs of one-particle states. At the given total pair momentum, such excitations belong to the continuous spectrum of the relative motion of pairs and manifest themselves as a sharp increase in the density of states in a narrow energy interval, on the order of the quasi-stationary state damping. Generally, quasi-stationary states are formed by all one-particle states from the region of definition of the relative pair motion momentum. It follows that the suppression of the density of one-particle states over the whole range of variations in the energy of relative motion corresponds to the formation of a quasi-stationary state. In this sense, the appearance of a pseudogap in the spectrum of one-particle excitations, which is observed in cuprates, can be directly related to the formation of quasi-stationary states.

We stress that quasi-stationary states, like the stationary state of a pair of particles or holes, are formed as solutions to the same equation (18) for the wave function of the relative motion of a pair.

Inequality (60) leads us to conclude that quasi-stationary states and the pseudogap mode related to them exist in a limited temperature range. Indeed, the interaction energy between particles $V(\mathbf{k} - \mathbf{k}')$ enters into the equation of pair motion (18) with weight $\Theta(\mathbf{k}')$, which is the characteristic function (17) of the kinematically allowed region Ξ at $T = 0$, namely, $\Theta(\mathbf{k}') = 1$ inside Ξ and $\Theta(\mathbf{k}') = 0$ outside Ξ. Because of the temperature dependence of the mean occupation numbers, the contribution of the $\Theta(\mathbf{k}')$ function to the integral in (18) decreases as temperature increases, which can be interpreted (as has already been mentioned above) as temperature weakening of intercomponent interactions in pairs. The corresponding decrease in the effective coupling constant U_0 results in the absence of a solution to (59) and, therefore, nonexistence of quasi-stationary states with low damping starting with the $U_0^{(m)}$ value corresponding to a certain temperature T^*. The T^* temperature takes on different but close values for quasi-stationary states with different total momenta close to optimal; therefore, the transition (crossover) from the pseudogap state to the state of a normal Fermi liquid should be observed in some temperature interval.

The crossover temperature T^* has the same energy scale as the superconducting transition temperature T_C. Indeed, if we suggest that the effective coupling constant has some U_0 value from interval (60) at a given level of doping and $T = 0$, then we must expect it to decrease as temperature increases because of temperature-induced weakening of interaction. As the pairs become noncoherent at T_C, the temperature dependence of the effective coupling constant can be estimated as $U_0(T) = U_0 \tanh(T_C/T)$. The crossover temperature can therefore be estimated as $T^* = T_C \tanh^{-1}(U_0^{(m)}/U_0)$.

The presence of quasi-stationary states as elementary excitations of the electronic system, which manifests itself in the temperature interval $T_C < T < T^*$, can cause this system to be substantially different from the normal Fermi liquid above the superconducting transition temperature. The phenomenological scheme [35] developed to describe the state of the quasi-two-dimensional electronic system formed from the superconducting condensate of pairs as the temperature increases ("algebraic Fermi liquid" [36, 30], whose properties are in some sense intermediate between the properties of the normal Fermi liquid and the Luttinger liquid, in which collective excitations in the form of charge and spin density waves are well defined and quasi-particle excitations strongly decay) presupposes that the transition from the superconducting to the normal state occurs as a Berezinski–Kosterlitz–Thouless transition as a result of phase disturbance of the wave function of the condensate caused by the rupture of "vortex–antivortex" pairs. The theory suggested in [35], which is asymptotically exact in the $1/N$ parameter, where N is the number of nodes of the superconducting order parameter, therefore presupposes the existence of noncoherent states of pairs in the temperature interval $T_C < T < T^*$; it is capable of explaining many properties of the pseudogap states of nondoped cuprates. Note that, when pairing occurs with a large total momentum and repulsive interaction, the number of nodes (the number of the intersection points between the line of pair wave function zeros and the Fermi contour) is $N \geq 8$. In addition, the line of zeros can itself be close to the Fourier contour regions that form the pair Fourier contour. Formally, the approach of the line of zeros to the pair Fermi contour means the passage to the $N \longrightarrow \infty$ limit, which broadens the applicability range of the theory suggested in [35]. Simultaneously, this approach increases the contribution of gapless excitations that arise at every node and favor the destruction of superconductivity. This causes exponential lowering of the amplitude of the superconducting order parameter [13].

8. CONCLUSIONS

The wave function $\Theta(\mathbf{k})\psi(\mathbf{k})$ of the relative motion of a pair [Eq. (23)] in the bound state is defined in region Ξ of the momentum space with characteristic dimensions Δk_1 and Δk_2 in the directions of the coordinate axes k_1 and k_2. The characteristic pair dimensions along the corresponding coordinate axes x_1 and x_2 can be estimated as $\Delta x_1 \sim (\Delta k_1)^{-1}$ and $\Delta x_2 \sim (\Delta k_2)^{-1}$, and some mean pair size, as $\xi_0 \sim 1/\sqrt{\Xi}$. Such estimates can be made for both the stationary and quasi-stationary pair states [5]. As directly follows from (18), the wave function of a bound stationary state is not alternating in region Ξ. Indeed, the bound state energy $E < 0$; therefore, the Green function in (18) is negative definite, $G(\mathbf{k}; E) < 0$. For this reason, $\psi(\mathbf{k})$ should necessarily change sign on some line inside Ξ [as follows from (23), this line is an arc of a circle if kernel (28) is degenerate) to ensure the equality of the left- and right-hand sides of (18) at $U(\mathbf{k} - \mathbf{k}') > 0$. Such lines of zeros of the wave function of relative motion exist both in the $\Xi^{(+)}$ region, where a pair of particles is excited, and in the $\Xi^{(-)}$ region, where a pair of holes is excited. The wave function of the relative motion of a quasi-stationary state with a low excitation energy also has a line of zeros, because otherwise, the integral over Ξ of the product of the real part of the Green function $G(\mathbf{k}; E)$ and the positive function $U(\mathbf{k} - \mathbf{k}')$ would be an essentially negative value. Note that the superconducting order parameter that appears in pairing with a large total momentum and repulsive interaction [13] also has a line of zeros that intersects the pair Fermi contour. The independently determined lines of zeros in the $\Xi^{(-)}$ (for pairs of holes) and $\Xi^{(+)}$ (for pairs of particles) regions transform exactly into this line when scattering between $\Xi^{(-)}$ and $\Xi^{(+)}$ is "switched on."

The superconducting order parameter that describes the condensate of pairs with the total momentum $\mathbf{K}$ therefore twice vanishes inside the Ξ_K region. Because of degeneracy along the directions crystallographically equivalent to the direction of the $\mathbf{K}$ vector, the order parameter that corresponds to the currentless state of the condensate is represented by a linear combination of crystallographically equivalent functions [5], each with two nodes on the Fourier contour. It follows that the superconducting order parameter of a tetragonal crystal, which transforms under the trivial A_{1g} irreducible representation, does not change sign under rotation through $\pi/2$ but has eight nodes on the Fermi contour and can be assigned extended s-wave symmetry or $(s + g)$ symmetry [37]. If the order parameter transforms under another one-dimensional irreducible representation (B_{1g}), it has twelve nodes on the Fermi contour and can formally be assigned d-wave (or $d + g$) symmetry, because it changes sign under rotation through $\pi/2$.

Order parameter symmetry is determined by the interaction that mixes states in the Ξ_K regions where the relative motion momenta of pairs with different crystallographically equivalent total momenta are defined. Coulomb repulsion (weaker because of substantial momentum transfer compared with the scattering of pairs within each of the Ξ_K regions), clearly, results in A_{1g} symmetry. Interaction caused by antiferromagnetic fluctuations [38] mixes states corresponding to pair momenta rotated through $\pi/2$ with respect to each other. If this interaction prevails over Coulomb scattering between crystallographically equivalent Ξ_K regions, it results in B_{1g} order parameter symmetry. Note that the role played by antiferromagnetic fluctuations increases as the level of doping of cuprates decreases, which does not exclude the possibility of the $A_{1g} \longleftrightarrow B_{1g}$ change in order parameter symmetry at a certain doping level.

The dependence of the superconducting order parameter on the pair kinetic energy $\xi(\mathbf{k})$ of the form $\Delta = a + b\xi$, which was considered in [39], is naturally related to asymmetry of current–voltage characteristics observed in cuprates. For this reason, the mere "tilt" of the superconducting gap (the ratio of the gap parameters $a/b \neq 0$) can serve as a basis [39] for selecting both interactions responsible for superconductivity and models for describing the superconductivity of cuprates. In this sense, pairing with a large total momentum and repulsive interaction, which results in the appearance of a line of zeros [13] and, therefore, some mean gap tilt, is compatible with the observed asymmetry of tunnel characteristics.

The bound state energy (50) exponentially depends on the density of states $g(0)$ of relative motion on the pair Fourier contour. If region Ξ has a hyperbolic metric (the metric is hyperbolic if the Fermi contour is situated in an extended neighborhood of the saddle point of the electronic dispersion law), the $g(0)$ value depends on how close is the logarithmic Van Hove singularity of the density of states of relative pair motion to the pair Fermi contour, $g(0) \equiv g(0; \zeta_s)$, where ζ_s is the energy of the logarithmic singularity counted from the pair Fermi contour (Fig. 2b). We can therefore assert [40] that the closeness of the Fermi contour to the saddle point is the reason for an increase in the density of states on the Fermi contour, $g(0; \zeta_s) \propto g(0)\ln\zeta_0/\zeta_s$, and the corresponding increase in the bound state energy given by (50). Note that the logarithmic singularity degenerates into the root singularity $g(0; \zeta_s) \propto E_s^{-1/2}$ in the case of an extremely strong anisotropy of effective masses [14]. As follows from (57), the closeness of the Fermi contour to the saddle point is also the reason for an increase in the energy and damping of quasi-stationary states.

ACKNOWLEDGMENTS

This work was financially supported by the Ministry of Education of Russia (project no. E02-3.4-147), the Federal Program "Integratsiya" (project no. B0049), the Russian Foundation for Basic Research (project no. 02-02-17133), and the Federal Scientific-Technical program "Studies and Developments in Priority Directions of Science and Technology" (state contract nos. 40.072.1.1.1173 and 40.012.1.1.1357).

REFERENCES

1. L. N. Cooper, Phys. Rev. **104**, 1189 (1956).
2. J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **108**, 1175 (1957).
3. P. Fulde and R. A. Ferrel, Phys. Rev. **135**, A550 (1964).
4. A. I. Larkin and Yu. N. Ovchinnikov, Zh. Éksp. Teor. Fiz. **47**, 1136 (1964) [Sov. Phys. JETP **20**, 762 (1964)].
5. V. I. Belyavskiĭ, V. V. Kapaev, and Yu. V. Kopaev, Zh. Éksp. Teor. Fiz. **118**, 941 (2000) [JETP **91**, 817 (2000)].
6. A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskiĭ, *Methods of Quantum Field Theory in Statistical Physics* (Fizmatgiz, Moscow, 1962; Prentice Hall, Englewood Cliffs, N.J., 1963).
7. P. W. Anderson, Science **256**, 1526 (1992).
8. J. Orenstein and A. J. Millis, Science **288**, 468 (2000).
9. Z.-X. Shen, W. E. Spicer, D. M. King, *et al.*, Science **267**, 343 (1995).
10. V. I. Belyavskiĭ and Yu. V. Kopaev, Zh. Éksp. Teor. Fiz. **121**, 175 (2002) [JETP **94**, 149 (2002)].
11. V. I. Belyavsky and Yu. V. Kopaev, Phys. Rev. B **67**, 024513 (2003).
12. V. I. Belyavskiĭ, V. V. Kapaev, and Yu. V. Kopaev, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 51 (2002) [JETP Lett. **76**, 44 (2002)].
13. V. I. Belyavskiĭ, Yu. V. Kopaev, V. M. Sofronov, and S. V. Shevtsov, Zh. Éksp. Teor. Fiz. **124**, 1149 (2003) [JETP **97**, 1032 (2003)].
14. A. A. Abrikosov, Physica C (Amsterdam) **341**–**348**, 97 (2000).
15. E. G. Maksimov, Usp. Fiz. Nauk **170**, 1033 (2000) [Phys. Usp. **43**, 965 (2000)].
16. H. Aoki, cond-mat/0305490.
17. P. W. Anderson, Science **237**, 1196 (1987).
18. G. I. Japaridze, A. P. Kampf, M. Sekania, *et al.*, Phys. Rev. B **65**, 014518 (2002).
19. J. C. Campuzano, M. R. Norman, and M. Randeria, in *Physics of Conventional and Unconventional Superconductors*, Ed. by K. H. Bennemann and J. B. Ketterson (Springer, Berlin, 2002).
20. A. Damascelli, Z. Hussain, and Z.-X. Chen, Rev. Mod. Phys. **75**, 473 (2003).
21. D. Pines and P. Noziéres, *Theory of Quantum Liquids* (Benjamin, New York, 1966; Mir, Moscow, 1967).
22. A. B. Vasil'eva and N. A. Tikhonov, *Integral Equations* (Nauka, Moscow, 2002) [in Russian].
23. G. M. Fikhtengol'ts, *Course of Differential and Integral Calculus* (Fizmatgiz, Moscow, 1962), Vol. 2 [in Russian].
24. A. A. Abrikosov, *Fundamentals of the Theory of Metals* (Nauka, Moscow, 1987; North-Holland, Amsterdam, 1988).
25. J. R. Schrieffer, *Theory of Superconductivity* (Benjamin, New York, 1964; Nauka, Moscow, 1970).
26. A. I. Baz', Ya. B. Zel'dovich, and A. M. Perelomov, *Scattering, Reactions and Decays in Nonrelativistic Quantum Mechanics*, 2nd ed. (Nauka, Moscow, 1971; Israel Program for Scientific Translations, Jerusalem, 1966).
27. E. M. Lifshitz and L. P. Pitaevskiĭ, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics*, 2nd ed. (Nauka, Moscow, 2001; Pergamon, New York, 1980).

28. V. J. Emery and S. A. Kivelson, Nature **374**, 434 (1995); Phys. Rev. Lett. **74**, 3253 (1995).
29. L. B. Ioffe and A. J. Millis, Science **285**, 1241 (1999).
30. M. Franz and Z. Tešanović, Phys. Rev. Lett. **87**, 257003 (2001).
31. J. B. Marston and I. Affleck, Phys. Rev. B **39**, 11538 (1989); **61**, 14773 (2000).
32. S. Chakravarty, R. B. Laughlin, D. K. Morr, and C. Nayak, Phys. Rev. B **63**, 094503 (2001).
33. D. A. Ivanov, P. A. Lee, and X.-G. Wen, Phys. Rev. Lett. **84**, 3958 (2000).
34. S. Chakravarty, Phys. Rev. B **66**, 224505 (2002).
35. M. Franz, Z. Tešanović, and O. Vafek, Phys. Rev. B **66**, 054535 (2002).
36. L. Balents, M. P. A. Fisher, and C. Nayak, Phys. Rev. B **60**, 1654 (1999).
37. B. H. Brandow, Phys. Rev. B **65**, 054503 (2002).
38. P. Monthoux, A. V. Balatsky, and D. Pines, Phys. Rev. Lett. **67**, 3448 (1991).
39. J. E. Hirsch, Phys. Rev. B **59**, 11962 (1999).
40. V. I. Belyavskiĭ and Yu. V. Kopaev, Pis'ma Zh. Éksp. Teor. Fiz. **72**, 734 (2000) [JETP Lett. **72**, 511 (2000)].

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 3, 2004, pp. 602–612.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 3, 2004, pp. 691–703.

SOLIDS
Electronic Properties

Chirality of a Forming Spin Spring and Remagnetization Features of a Bilayer Ferromagnetic System

V. S. Gornakov[a], Yu. P. Kabanov[a], V. I. Nikitenko[a], O. A. Tikhomirov[a], A. J. Shapiro[b], and R. D. Shull[b]

[a]*Institute of Solid State Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*
e-mail: gornakov@issp.ac.ru
[b]*National Institute of Standards and Technology, MD 20899, Gaithersburg, USA*

Received February 17, 2004

Abstract—Distribution of a magnetic moment in an exchange-coupled bilayer Fe/SmCo epitaxial structure grown on a (110) MgO substrate is visualized by the magnetooptic indicator film technique. The direction and the magnitude of the effective magnetization in this structure are determined both under external magnetic fields of variable magnitude and direction and after the removal of these fields. It is shown that such a heterostructure is remagnetized by a nonuniform rotation of a magnetic moment both along the thickness of a sample and in its plane. A field antiparallel to the axis of unidirectional anisotropy gives rise to spin springs with opposite chiralities in different regions of the magnetically soft ferromagnetic layer. The contributions of these springs to the net magnetization cancel out, thus decreasing the averaged magnetic moment and the remanent magnetization without their rotation. When the external field deviates from the easy axis, the balance is violated and the sample exhibits a quasi-uniform rotation of the magnetic moment. Asymmetry in the rotation of the magnetic moment is observed under the reversal of the field as well as under repeated remagnetization cycles. It is established that a monochiral spin spring is also formed in a rotating in-plane magnetic field when the magnitude of the field exceeds the critical value. Possible mechanisms of remagnetization in this system are discussed with regard to the original disordered orientation of magnetization of the magnetically soft layer with respect to the easy axis, which is defined by the variance of unidirectional anisotropy axes of this layer on the interface. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The development of layered nanocomposite materials has opened a new chapter in the physics of magnetism [1]. The exchange interaction on the interface between layers with different magnetic ordering forms an essentially new ground state of a heterophase magnet, radically changes the behavior of spins in an external magnetic field, and gives rise to a number of new and unusual phenomena [2–7]. These include the phenomenon of unidirectional (exchange) anisotropy, which manifests itself in the shift of hysteresis loops along the magnetic-field axis with respect to the origin of coordinates, in the considerable increase in the coercivity of the ferromagnet, in the anomalous rotational hysteresis of a bilayer structure, etc. The exchange-induced shift of a hysteresis loop may reach several tens of oersteds and has already been used in magnetic-field sensors based on the giant magnetoresistance phenomenon as well as in new computer memory elements.

In spite of the fact that the properties of bilayer nanocomposite structures consisting of exchange-coupled magnetically hard and soft ferromagnetic layers have been intensively studied, a number of fundamental features of the remagnetization of these structures remain unclear. Currently, the simplest mechanism, associated with the formation of a one-dimensional heterophase spin spiral (exchange spring) in an external magnetic field, is being extensively discussed. In such a spring, the spins of adjacent atomic layers in the magnetically soft layer smoothly turn, under an external field, from the direction of the effective field on the surface to the direction of the unidirectional anisotropy field near the interface [8–11]. Calculations on the basis of this idealized model give only a qualitative explanation to certain observable remagnetization features, such as an exchange-induced shift of a minor hysteresis loop of the heterostructure, critical fields needed for the beginning of the formation of a spin spiral, and the subsequent reversibility of the initial stage of remagnetization.

When analyzing the behavior of real layered nanocomposites, one should take into account that the equivalence of the energy of the exchange spin spring with respect to its twisting direction may give rise to domains with opposite rotation of spins in different regions of the film. A real crystal contains magnetostatic fields, lattice defects, boundaries of blocks and

grains, and steps on the interface. All these factors lead to the misorientation of the anisotropy fields and may stimulate a rotation of spins in opposite directions.

Similar two-dimensional spin spirals, but topologically stable ones—domain walls—are formed in thin films and bulk samples of homogeneous ferromagnets [12, 13]. Usually, domain walls consist of subdomains that are characterized by different chiralities and are separated by Bloch lines, which, in turn, may be separated by Bloch points. The formation and the dynamic properties of two- and three-dimensional walls in these materials have been studied in sufficient detail both theoretically and experimentally [13–15].

The analysis of similar phenomena that occur in thin-film heterophase structures is presently at its initial stage. However, it is already clear that the investigation of these phenomena is a necessary step toward an adequate description of microscopic mechanisms of remagnetization in layered magnetic nanocomposites. Another important feature is the fact that the thickness of a magnetically soft layer in typical structures consisting of magnetically soft and hard layers is no greater (and usually much less) than the width of a Bloch wall $\delta \sim \sqrt{A/K}$, where A and K are the exchange and crystallographic anisotropy energies. This means that the whole remagnetization process of a magnetically soft layer occurs only due to the nucleation and the development of a partial domain wall (spin spiral or exchange spring) parallel to the surface of the film rather than due to its translational motion as in single-phase materials. Moreover, a full switching of magnetization in a magnetically soft layer (and in the entire heterostructure) should strongly depend on the interaction between the forming exchange spring and the spin subsystems localized both in the bulk of the magnetically hard layer and near the interface.

The boundaries between regions with different chiralities in a quasi-two-dimensional heterophase exchange spring are somewhat similar to the well-known Bloch lines in domain walls, but they should be characterized by a specific nontrivial spin structure because they are bounded by interphase surfaces rather than by domains. By now, a direct experimental study of the formation and the development of a spin spiral (which constitutes a partial domain wall) under an external magnetic field has been performed in a bilayer thin-film heterostructure [16] by the magnetooptic indicator film technique [17]. In [16, 17], the initial stage of remagnetization of a layered nanocomposite consisting of a thin (50 nm) magnetically soft iron layer and a high-coercivity 35-nm-thick SmCo layer grown on the (100) surface of single-crystal MgO was studied. It was found that, despite theoretical predictions, a one-dimensional exchange spring (or a spin spiral) was not formed in a field that is strictly antiparallel to the macroscopic unidirectional anisotropy axis. The system was remagnetized by a nonuniform rotation of spins that gave rise to a two-dimensional spin spiral, while the magnetic moment **M**, averaged over the thickness of the heterostructure, retained its direction along the unidirectional anisotropy field. The quasi-one-dimensional spin spiral was only formed when the magnetic field was tilted to the easy axis. It was the only case when the evolution of the spring under increasing field was accompanied by the rotation of the vector **M**.

In [18], the authors investigated the same Fe/SmCo system as that in [16, 17] but with somewhat different thicknesses of the layers and grown on a different plane of single-crystal MgO, namely, on the (110) plane. The authors of [18] observed a rotation of the vector **M** both during the remagnetization of a single magnetically soft layer and after the removal of a large external magnetic field that stimulated the penetration of the exchange spring from the magnetically soft layer into the high-coercivity layer. The last effect, the deviation of the averaged remanent magnetization from the initial direction, parallel to the unidirectional anisotropy axis, was not previously predicted. It was interpreted by the authors of [18] as a result of competition between the intrinsic and the interlayer exchange interactions that gave rise to an apparent biquadratic interaction, similar to that in the Slonczewski model [19]. It should be emphasized that, in [18], the rotation of averaged magnetization, which was evidence of the formation of a one-dimensional exchange spring, was observed in a field that was nominally antiparallel to the anisotropy axis, in contrast to the results presented in [17].

To determine the conditions and the formation mechanisms of the phenomena revealed in [16–18] and to implement one of possible switching modes associated with the formation of spin spirals of different dimensions, in the present work, we carry out detailed investigations of the evolution of the net magnetic moment in the structure consisting of magnetically soft and hard layers as a function of the angle between the external field and the macroscopic unidirectional anisotropy axis under repeated remagnetization cycles. As a result, we show that an inhomogeneous two-dimensional spin spiral is also formed in a Fe/SmCo nanocomposite grown on the (110) MgO substrate in a field antiparallel to the undirectional anisotropy axis; moreover, it is formed in a very narrow (less than one degree) interval of orientations of these fields. We also observe a new phenomenon of asymmetry in the development of the exchange spring under the reversal of the field during repeated remagnetization cycles.

2. EXPERIMENTAL TECHNIQUES

The sample under investigation was a Fe/SmCo/Cr epitaxial structure deposited by the magnetron sputtering technique on a single crystal (110) MgO substrate. The thicknesses of the magnetically hard layer, magnetically soft layer, and the chromium buffer layer were 35 nm, 50 nm, and 20 nm, respectively. More detailed description of the fabrication technology is given in [11].

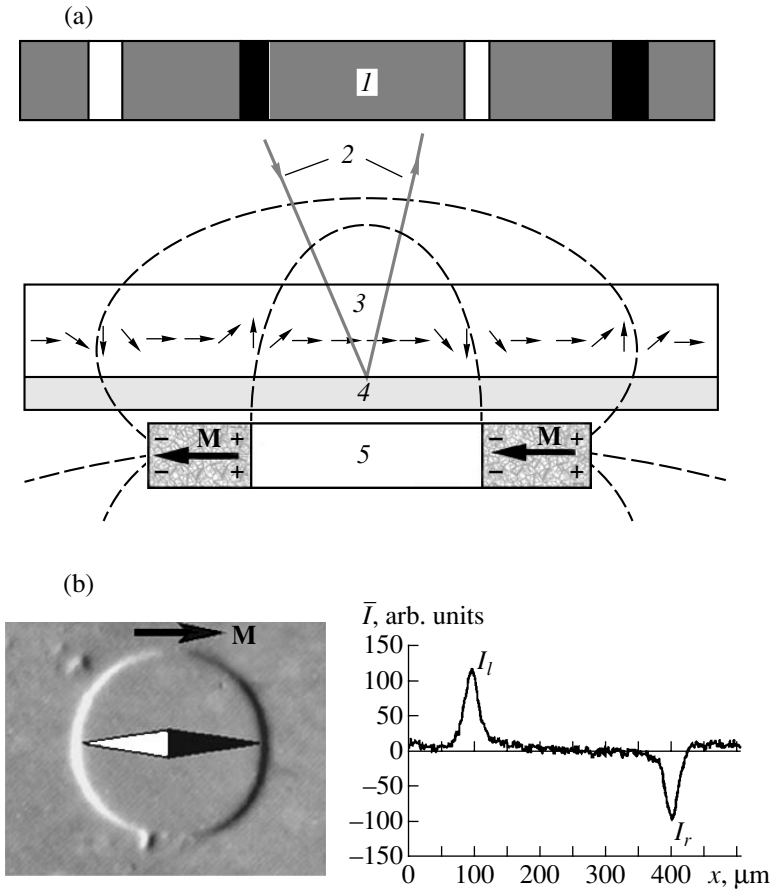

Fig. 1. Determination of the direction of magnetization by a magnetooptic indicator film. (a) Scheme of the experiment and magnetooptic contrast due to stray fields: (*1*) distribution of magnetooptic contrast, (*2*) linearly polarized light, (*3*) indicator film, (*4*) deposited mirror film of aluminum, and (*5*) sample with a hole. (b) Magnetooptic portrait of magnetization near the edge of the hole and the intensity profile of the magnetooptic signal.

Remagnetization processes were studied by visualizing stray magnetic fields with the use of a magnetooptic indicator film [17, 20] that was placed immediately on the surface of the sample (Fig. 1a). In the absence of external magnetic fields, the magnetic moment of the indicator film lies in the plane of the film; however, under the normal component of the stray fields of the sample, the local magnetization of the indicator film deviates from the plane, thus rotating the polarization plane of light due to the Faraday effect. Polarized light is incident normal to the indicator film and is reflected from the surface adjoining the sample, which is coated with a thin aluminum film to improve the reflectivity. When the polarizer and analyzer are slightly uncrossed, a magnetooptic pattern looks either dark or light, with appropriate variations of intensity, depending on the magnitude and the sign of the local stray field.

To analyze the magnetization in the plane of the sample, a round through hole with a diameter of 300 μm was made in the latter (Fig. 1b). When the sample is magnetized uniformly, the direction of the magnetization **M** is determined by the orientation of the symmetry axis (which is indicated by a compass needle) of the magnetooptic portrait formed by the components of the stray field at the edge of the round hole. It should be noted that the magnetostatic field around the probing hole, just as the stray fields associated with the edges of the film, did not give rise to closure domains and did not produce any appreciable effect on the magnetization of the film. A quantitative characteristic of magnetization is given by the averaged intensity of the magnetooptic signal,

$$\bar{I} = (I_l + I_r)/2,$$

where I_l and I_r are the maximal intensities on the dark and light edges, respectively. This characteristic is determined by the angle of Faraday rotation of the polarization plane of light, which is proportional to the perpendicular component of the stray field at the edge of the hole and, hence, to the in-plane magnetization averaged over the sample thickness.

In large external fields H ($\mu_0 H > 100$ mT), direct observation of the magnetic structure is complicated by a strong field that bends the magnetic moment of the indicator to the plane of the film. In this case, one

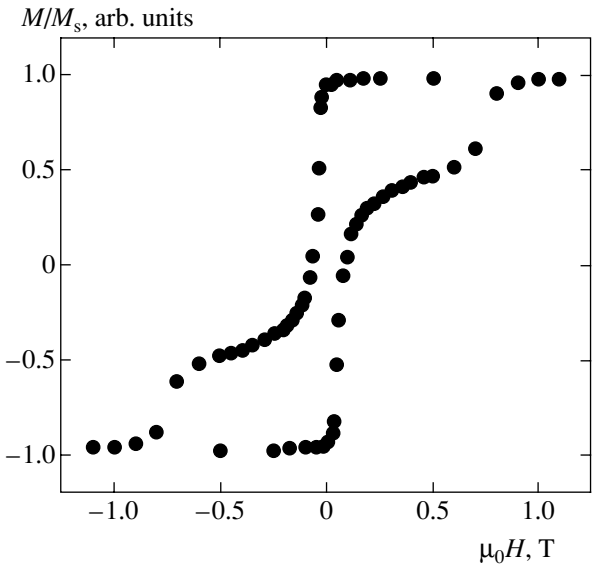

Fig. 2. Hysteresis loop of a Fe/SmCo sample obtained by a SQUID magnetometer.

observes a magnetooptic portrait of remanent magnetization after the application and the removal of the field of necessary magnitude.

3. EXPERIMENTAL RESULTS AND THEIR DISCUSSION

Figure 2 shows a hysteresis loop of the bilayer system under investigation that was obtained by a SQUID magnetometer when the field was applied along the easy axis of the magnetically hard Sm_2Co_7 film. One can easily see two characteristic stages of the remagnetization process: the first is largely associated with the magnetically soft iron layer (up to about 150 mT), and the second, with the magnetically hard SmCo layer.

The specific features of the magnetization process revealed by a magnetooptic indicator film are shown in Fig. 3. In the initial state, the directions of magnetizations in the magnetically soft and hard layers coincide due to the unidirectional exchange anisotropy induced by the SmCo layer. A magnetooptic signal proportional

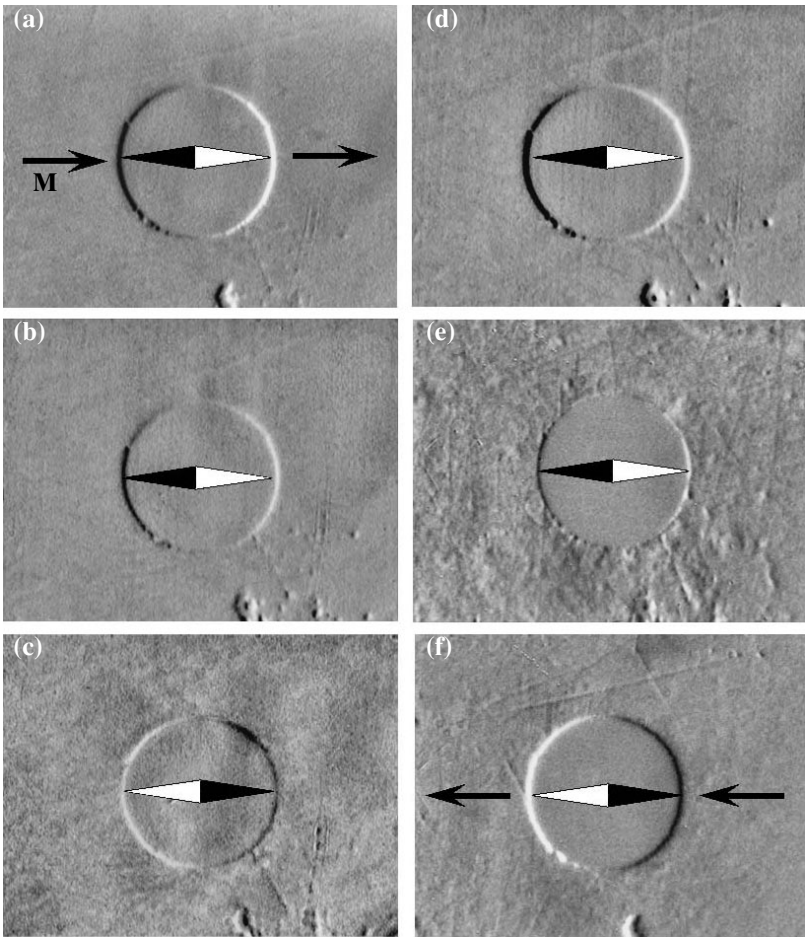

Fig. 3 Evolution of the magnetooptic portrait of the Fe/SmCo system under its (a–d) reversible and (e–f) irreversible remagnetization along easy axis: (a) initial state ($H = 0$) after the removal of a large positive field of $\mu_0 H = +7.0$ T; (b) and (c) the same for $\mu_0 H =$ –0,02 and –0.09 T, respectively; (d), (e), and (f) after the removal of a field of $\mu_0 H = –0.09$, –0.75, and –0.87 T, respectively.

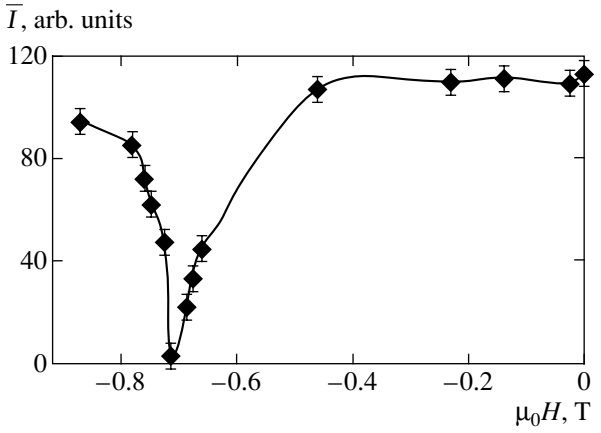

Fig. 4. Variation of the averaged maximal intensity $\overline{I}$ of the remanent magnetooptic signal at the edge of the hole under remagnetization along easy axis.

to the magnitude of the stray field vanishes at the point at which the direction of magnetization coincides with the tangent to the edge of the hole (Fig. 3a). Along the horizontal axis, where the radial component of magnetization and the density of induced magnetic charges at the edge of the hole are maximal, one observes a strong magnetooptic signal; the signs of this signal on the right and left edges are opposite (positive and negative magnetic charges). The observed pattern of the stray fields reflects the two-dimensional distribution of magnetization averaged over the thickness.

Application of the magnetic field of opposite direction changes the distribution of magnetization. This fact manifests itself in the reduced magnetooptic contrast at the edges of the hole (Fig. 3b). A further increase in the field gives rise to spatial fluctuations of the magnetooptic signal that correspond to a nonuniform distribution of magnetization along the sample surface. The inversion of the magnetooptic contrast corresponds to the reversal of the magnetization (Fig. 3c). If the applied field is relatively small, then the remagnetization is completely reversible, and the initial pattern is fully recovered after the removal of the field: both layers are magnetized along the easy axis (Fig. 3d). In stronger fields, the reversibility is lost, and the magnetooptic contrast is not recovered after the removal of the field. Nevertheless, the direction of the remanent magnetization, which is determined by the symmetry axis of the magnetooptic pattern at the edges of the hole, remain unchanged and coincide with the direction of the easy axis (Fig. 3e). Under a further increase of the magnetizing field, the remanent magnetooptic contrast vanishes and then appears again, but with the opposite sign. Figure 3f clearly shows that, in the maximal field attainable in the experiment, magnetooptic contrast virtually reaches its initial absolute value.

In addition to the visual observation of the qualitative features of remagnetization, the magnetooptic indicator film technique allows one to measure the numerical characteristics of the remagnetization. Figure 4 represents the variation of the averaged intensity $\overline{I}$ corresponding to the variation of the remanent magnetization of the structure due to the magnetic field applied along the easy axis. The intensity of the magnetooptic signal is obtained by the digital processing of the stray field patterns, similar to those shown in Fig. 3, near the edge of the hole. One can see that, up to the fields of 0.5 T, the variations of magnetization are fully reversible (the remanent magnetization is close to the initial value); a rapid decrease (virtually up to zero) of the magnetooptic signal, accompanied by the inversion of the contrast under a further increase in the field, occurs only in the neighborhood of 0.7 T. However, it is important that the initial intensity is not fully recovered up to a field of 0.9 T, and, hence, the structure is not magnetized up to saturation.

The observed variation of the magnetooptic portrait in a field directed along the easy axis corresponds to the remagnetization of the system by a nonuniform rotation of magnetization in submicron-size regions (beyond the spatial resolution limit of optical microscopy). A similar mode of remagnetization was earlier observed in Fe/SmCo bilayer structures with the (100) orientation of a MgO substrate [17].

The remagnetization kinetics described above is in an obvious contradiction with the results of [18], where the authors observed a uniform rotation of remanent magnetization in bilayer structures of identical composition but with slightly different structural and geometrical characteristics. We have established that a transition to a quasi-uniform remagnetization mode, which corresponds to the formation of a macroscopic exchange spring in a sample, occurs under a small (of about few fractions of a degree) deviation of the applied field from the easy magnetization axis. This remagnetization mechanism is illustrated in Fig. 5. The field is directed at an angle of $\alpha = -3°$ to the easy axis. Instead of successive vanishing and inversion of magnetooptic contrast, which was observed in the experiments in which the field is antiparallel to the unidirectional anisotropy axis, here we have a rotation of magnetization in the plane of the film. Just as in the case of a strictly antiparallel field, the first stage of remagnetization is reversible, and the magnitude and the direction of the magnetization are restored after the removal of the field. As the field increases, the resultant vector of remanent magnetization also begins to deviate from the easy axis (Figs. 5b–5e). The rotation is accompanied by a slight reduction of the magnetooptic contrast, which provides evidence for the inhomogeneity of this process. However, the contrast is not reduced to zero, as is the case in a strictly anisotropic field, and the direction of remanent magnetization is easily determined for any magnitude of the applied field.

The investigation of the full remagnetization cycle (under the reversal of the field) revealed another interesting phenomenon. The sign of rotation of the vector

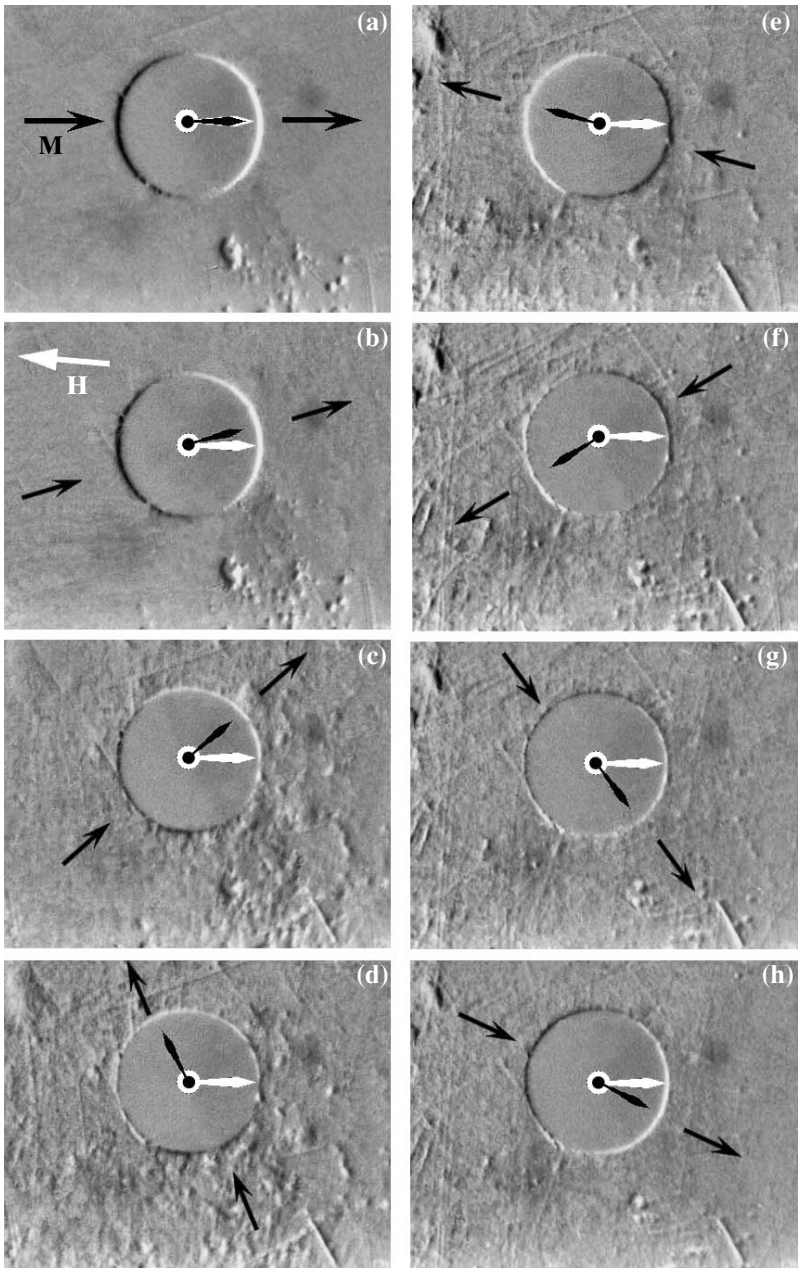

Fig. 5. Remagnetization of a Fe/SmCo heterostructure under small ($\alpha \approx -3°$) deviation of the field from easy axis; (a) in the initial state, $H = 0$; (b) $\mu_0 H = -0.57$ T; (c) $\mu_0 H = -0.72$ T; (d) $\mu_0 H = -0.78$ T; (e) $\mu_0 H = -0.87$ T; (f) $\mu_0 H = +0.35$ T; (g) $\mu_0 H = +0.48$ T; and (h) $\mu_0 H = +0.57$ T. The figure demonstrates the state of polarization after the removal of the field. Light and dark arrows inside the hole show the directions of the unidirectional anisotropy and of the net magnetization, respectively.

of remanent magnetization in the field of opposite direction sometimes coincided with the sign of rotation of the vector **M** on the previous branch of the hysteresis loop (Figs. 5f–5h). In other words, the angle of this rotation, which varied counterclockwise in the previous half-cycle of remagnetization, continued to vary in the same direction up to an angle of 360°, which coincides with the initial angle $\varphi = 0$. This sign of rotation remained unchanged during repeated remagnetization cycles. In some other cases, the variation of φ had opposite signs on the descending and ascending branches of the hysteresis loop; i.e., during reverse remagnetization, intermediate directions of remanent magnetization reproduced, in reverse order, the previous directions, while the values of the angle φ between 180° and 360° were not realized at all.

To investigate the characteristics of the asymmetry in the evolution and the sign of twisting of the spin spring, we performed a series of remagnetization cycles for different angles between the magnetic field and the

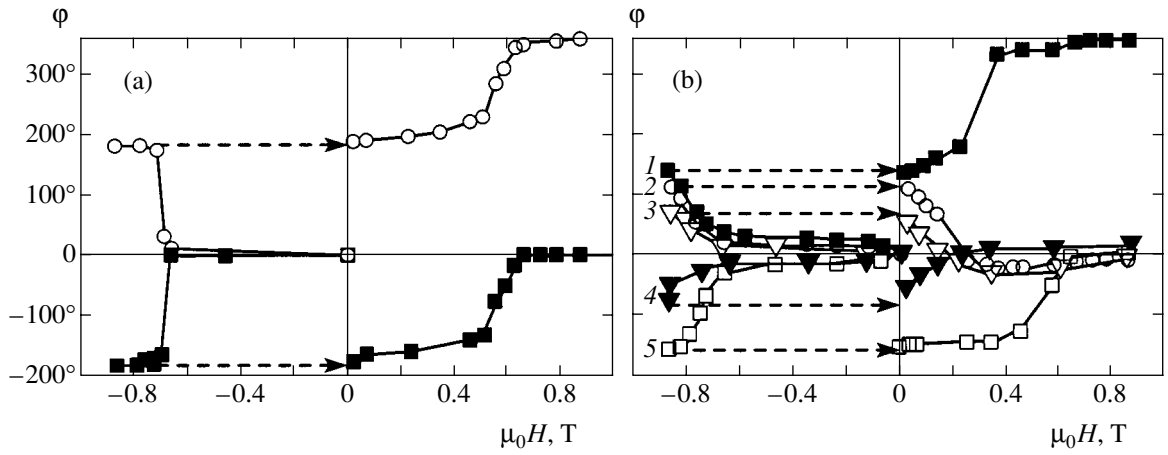

Fig. 6. Rotation of remanent magnetization under various tilt angles of a remagnetizing field with respect to easy axis; (a) small angles $\alpha = +0.5°$ and $\alpha = -0.5°$ and (b) large angles (*1*) $\alpha = -6°$, (*2*) –8°, (*3*) –10°, (*4*) 15°, and (*5*) 5°. Dashed curves with arrows correspond to transitions from a negative to a positive remagnetization half-cycle.

easy axis. Some of these results are shown in Fig. 6. It turned out that the rotation angle of the remanent magnetization is extremely sensitive to the sign and the relative magnitude of the magnetic field component perpendicular to the easy axis. For a nonuniform rotation of the remanent magnetization, which was observed when the field was strictly parallel to the easy axis, the angle φ is equal to zero for any value of the field. As described above, even small deviations (of about 1°) change the mode of remagnetization and result in a nonzero rotation of the remanent magnetization (Fig. 6a). For small tilt angles, the variation of φ is relatively small up to a certain critical value of the field at which this angle sharply increases. When the field increases further, this angle rapidly reaches its maximal value (which is close to 180°). When the sign of the tilt angle between the direction of the field and the easy axis is changed, the sign of rotation of the remanent magnetization is also reversed. The divergent curves in Fig. 6a are obtained for a relative rotation of the external field through an angle as small as 1°.

We have observed that an increase in the tilt angle of the field with respect to the easy axis substantially changes the evolution of the spin spring. For instance, a rotation of the field through only ±10° results in a decrease in the resultant angle of remanent magnetization from 180° to about 70° (curves *3* and *4* in Fig. 6b). The range of fields within which the angle φ shows significant variation is also changed: instead of a sharp increase of the angle to values close to 180° (Fig. 6a), the rotation of the remanent magnetization occurs slowly (curves *1*–*5*). Figure 6 illustrates the above-described asymmetry in the twisting sign of the spin spring under the reversal of the remagnetizing field. When the field was tilted with respect to the easy axis at a small angle in the counterclockwise direction, then, after the reversal of the field, the sign of rotation of the remanent magnetization **M** coincided with the sign of rotation at the preceding half-cycle of remagnetization (curve *1* in Fig. 6b). Thus, the angle φ varied from 180° to 360°. When the field was tilted in the opposite direction, then the sign of twisting of the spin spring was changed, and the angle φ ran through the same values as in the preceding half-cycle, but in the reverse order (curve *5*). Since the intervals of angles from 180° to 360° and from –180° to 0° are physically equivalent, we can conclude that there exists a distinguished sign of rotation of remanent magnetization when a sample is remagnetized in a slightly tilted field with the magnitude ranging from –0.9 T to 0.9 T. Figure 6b shows that, for large tilt angles of the field (greater than 7°), the variation of φ in the reverse half-cycle is always opposite to that in the initial half-cycle (curves *3* and *4*). Figure 6b also shows that this variant is always realized when the angle of rotation of remanent magnetization does not reach 120° in absolute value after the application of the maximal field of 0.9 T.

The remagnetization of multilayer magnetic systems exhibits a wide variety of mechanisms compared with the remagnetization of bulk materials. The main reason lies in the presence of additional factors—the interlayer exchange interaction and the unidirectional exchange anisotropy. When studying the remagnetization of similar systems, most authors restrict themselves to one-dimensional models and consider only spatial variations of magnetization that is perpendicular to the layers, assuming that the distribution along the remaining two coordinates is uniform [8–11]. However, it is well known from experience in the study of bulk samples that idealized mechanisms of uniform remagnetization very rarely occur in real samples and the key effect on the remagnetization of a specific system is produced by the irregularities of the crystalline structure and by magnetostatic fields.

One of the mechanisms by which irregularities affect the remagnetization of heterostructures with spin springs was considered in [18] while explaining the phenomenon of irreversible rotation of remanent mag-

netization. It was demonstrated that the key role in this process is played by the competition between interlayer exchange interaction and the exchange inside a soft layer. The presence of small-scale (comparable to the size of grains in a sample) domains of opposite polarities in the SmCo layer should form an appropriate "domain structure" from local spirals of opposite chiralities in the magnetically soft layer. On the other hand, such rapid spatial variations of magnetization in the iron layer would significantly increase both the exchange and the magnetostatic energy of the system. The competition between these two opposite trends may give rise to an intermediate state in which the magnetization of the magnetically soft layer is approximately uniform along two in-plane coordinates but is turned to a certain compromise angle with respect to the magnetization of the magnetically hard layer [18]. The rotation angle of the remanent magnetization is determined by the relative density of domain states in the SmCo layer; the domains of opposite polarities serve as local sources of nanosprings that twist the remanent magnetization of the soft layer in one or other direction. The rotation of the remanent moment of the magnetically soft layer due to the rearrangement of the domain structure of the SmCo layer is similar in appearance to the manifestation of a hypothetical biquadratic exchange interaction, whose presence in layered magnetic systems has been extensively discussed over the last decade [19, 21–23]. Indirect evidence of the fact that the observed rotation of the remanent magnetization is not associated with a certain additional fundamental interaction is the fact that the perpendicular configuration of magnetic moments of the layers, which is characteristic of the biquadratic exchange interaction, seems to be unstable and is virtually not observed in the experiment (see Fig. 5 and figures in [18]), whereas other intermediate orientations of spins are sufficiently widely presented.

The arguments presented in [18] account for the rotation mechanism of remanent magnetization under conditions when a quasi-uniform spin spring may be formed in a sample. Nevertheless, the results obtained in our studies show that such conditions are not always realized. Moreover, it turns out that the remagnetization of multilayer systems strongly depends on the initial conditions of the formation and evolution of a macroscopic spin spring.

One of the situations when the mechanism proposed in [18] does not work is the remagnetization in a field that is strictly parallel to the easy axis. The most probable factor that prevents the formation of a monochiral spin spring in a sample is the dispersion of the anisotropy axes associated with a columnar nanometer-grain structure that penetrates through all the layers of the structure. A weakly disordered orientation of the crystallographic axes in the grains, nonuniform stresses, the coexistence of different crystalline modifications of SmCo compounds, as well as the presence of steps on the interface [24] have been considered as mechanisms of such variance. When there is a magnetic field component perpendicular to the easy axis, this component determines a more favorable direction of twisting of spins, and the nuclei of spin springs in each grain are twisted in the same direction. Although this process is inevitably nonuniform due to the existence of a real potential relief, the presence of a strong exchange interaction smoothens these irregularities, thus forming a smooth distribution of spins over relatively large spatial areas. If there is no transverse component of the field, the direction of twisting in each nucleus of a spin spring is chosen randomly depending on the local potential microrelief. The interaction between these nuclei of different chiralities inevitably leads to a sharp increase both in the exchange energy and the energy of demagnetization fields, which is very large in a magnetically soft material. It is rather difficult to imagine the resultant magnetic configuration; however, it is most probable that this is a rippled three-dimensional vortex structure based on the balance of magnetostatic and exchange energies. In this case, elementary remagnetization processes occur on small spatial scales and cannot be resolved by optical microscopy (Fig. 3e).

Another moment that is not quite clear is the strong dependence of the rotation of remanent magnetization on the tilt angle of the magnetic field to the easy axis. One of possible explanations is the incompleteness of remagnetization processes in the SmCo layer even in the maximal field attainable in our experiment. However, it is unlikely that a small decrease in the longitudinal component of the field at a relatively small tilt angle may lead to such a sharp variation in the magnetization curves (see Fig. 6a). It is more probable that a reason should be sought for in the magnetically soft layer. Here, one should note that the twisting direction of spin springs is not yet quite obvious, the rotation of the magnetization vector **M** is in fact unstable, the process of remagnetization is avalanche-like, and a large part of intermediate angles is "skipped," leading directly to the opposite orientation. There may be a certain competition between rotation and 180-degree switching of magnetization; this is indirectly supported by the reduced magnetooptic contrast at the edges of the hole in the sample (see Fig. 4). The reasons for the "suppression" of the rotation of remanent magnetization for a large tilt angle (see Fig. 6c) are not quite clear at present. For example, this may be associated either with an increase in the magnetostatic energy or with a certain metastability of the remanent heterochiral structure under the variation of its spatial scale.

As for the asymmetry of rotation during a full remagnetization cycle, it also may be associated with the initial twisting direction of a spin spring. Suppose that a sample was initially remagnetized in a field directed to the left (Fig. 7a). In this case, the macroscopic spin spiral should be twisted so that the magnetization coincides with the transverse component of the

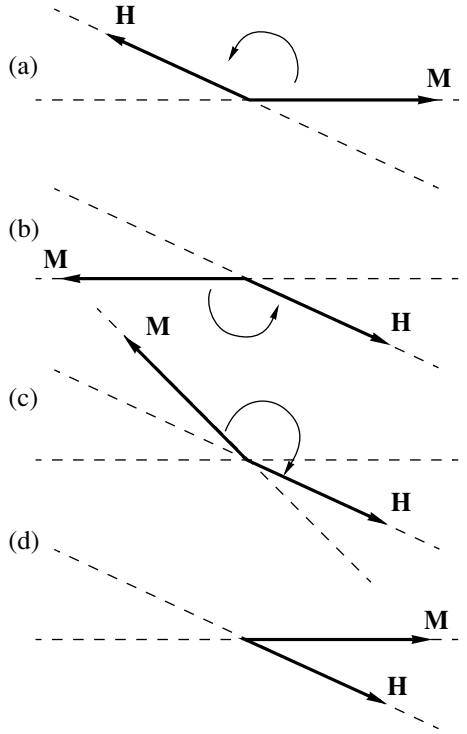

Fig. 7. Twisting direction of a spin spring under a full remagnetization cycle in a tilted field; (a) the first half-cycle of remagnetization; (b) effect of reverse field on remagnetized regions; (c) effect of a field on the inclined magnetization of the soft layer; (d) absence of a spin spring in macroscopic remanent domains.

field, as is shown in the figure. As a result, domains with the magnetization directed to the left are formed in the magnetically hard layer (Fig. 7b). After the removal of the field, the spins of the magnetically soft layer will be turned in the same direction, specified by the new direction of induced unidirectional anisotropy. Nevertheless, the presence of a nonremagnetized part in the bulk of the magnetically hard layer, which contains spins oriented in the previous direction (to the right; see Fig. 7d) gives rise to intermediate regions between the remagnetized and nonremagnetized parts of the layer, in which the magnetic moments of the magnetically soft layer are not completely reversed. As a result, the magnetization of the magnetically soft layer in these regions proves to be slightly tilted to the easy axis (Fig. 7c) and acts as localized nanosprings that tend to turn a part of spins in this layer in the previous direction.

When the external magnetic field is applied in the opposite direction, the twisting direction of a new spin spring is determined by the sign of the transverse projection of the field. It is obvious that, in the ideal case of total remagnetization, the rotation in the same direction is more favorable (Fig. 7b). However, due to the incompleteness of the remagnetization in the hard layer, the remanent magnetization after the first half-cycle often makes a considerable angle with the easy axis (see Fig. 6). If this angle is sufficiently large, then the sign of the transverse projection of the field proves to be opposite, and the spin spring starts to twist in the opposite direction (Fig. 7c). Thus, depending on the magnetization distribution in the crystal, three variants of the local effect of the magnetically hard layer on the magnetically soft one are possible under reverse remagnetization: an effect in the same direction, produced by the regions remagnetized during the previous half-cycle (Fig. 7b); an effect in the opposite direction, produced by the regions with tilted magnetization (Fig. 7c); and a nearly zero twisting moment produced by large nonremagnetized regions (Fig. 7d). The development of the resultant macroscopic spin spiral is determined by the statistical contribution of these three effects to the process of remagnetization. Since the remanent magnetization is maximally different from 180° at large tilt angles of the field (see Fig. 6), the sign of its rotation in a new half-cycle of remagnetization should always be reversed, which is observed in experiment.

Thus, if there is even a small factor in a sample that removes the degeneracy of states of the exchange spring with respect to its chirality, then the formation of a macroscopic spin spiral proves to be much more favorable compared with nonuniform rotation of the magnetic moment. This fact has been confirmed by experiments in which a rotating magnetic field of constant magnitude was used instead of remagnetization in a prescribed direction to form a spin spiral. The measurement of the longitudinal M_l and transverse M_t components of magnetization as a function of the rotation angle of the field shows a variation in the sign of twisting of the spiral when the deviation of the field from the easy axis is 15°–20° (see the jumps in M_t in Fig. 8). In this experiment, just as during the remagnetization along a fixed direction for $\varphi \neq 0$, the magnetooptic portrait illustrates the rotation of magnetization (Figs. 9a–9d); in this case, a decrease in intensity is not observed because the twisting moment is uniquely specified. The rotation angle of magnetization as a function of the current direction of the field is shown in Fig. 9e for several values of H. The following fact is noteworthy: a change of the sign of rotation of magnetization does not depend on the magnitude of the field and occurs soon after the field vector passes through the easy axis. This means that, during change of sign of the twisting moment (which is defined by the orientation of the field with respect to the magnetization of the hard layer), the spin spring relatively easily reverses its chirality. This process seems to be in a sense analogous to the generation of horizontal Bloch lines in twisted domain structures that occurs when the field in films with magnetic bubble domains reaches its critical value [13]. When the magnitude of the rotating field exceeds a certain critical value ($\mu_0 H > 60$ mT), the inhomogeneity of the remagnetization can be observed visually (Figs. 9c and 9d); in this case, a macroscopic domain structure is formed in the sample.

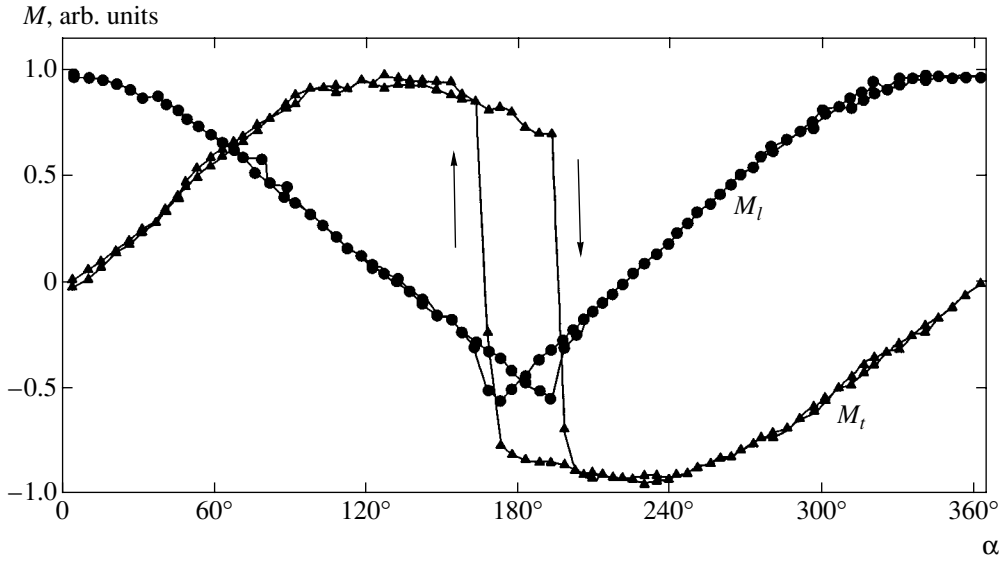

Fig. 8. Longitudinal and transverse components of magnetization in a rotating magnetic field of $\mu_0 H = 0.05$ T.

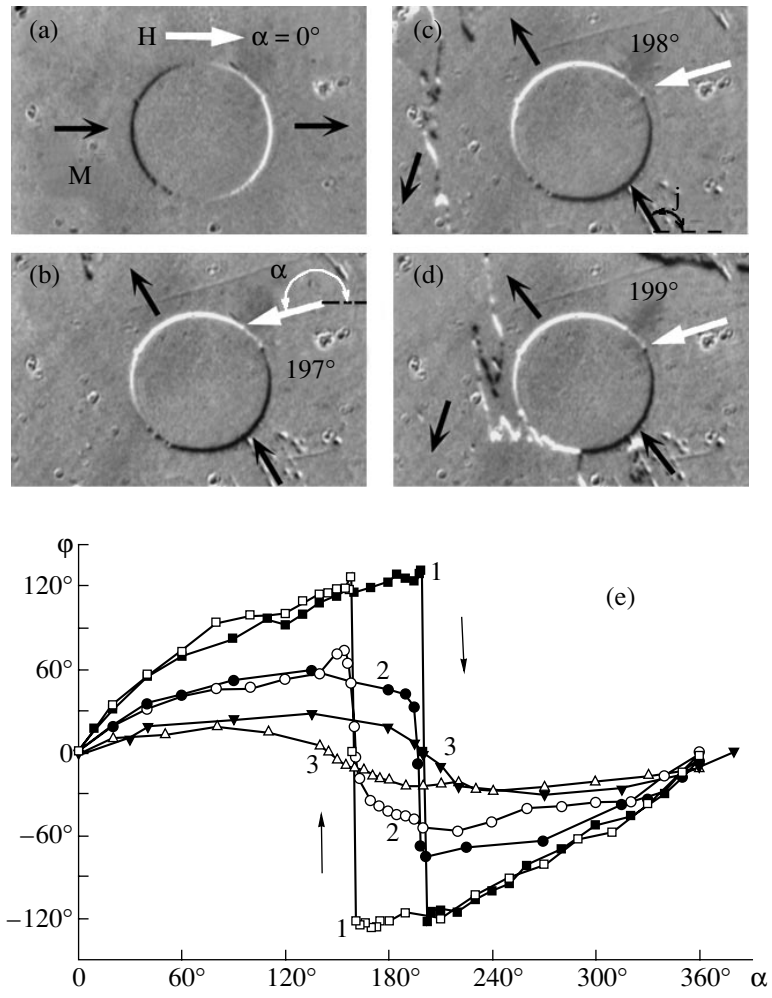

Fig. 9. Rotation of magnetization under rotating magnetic fields of various intensities; (a)–(d) successive stages of magnetization rotation and (e) angle of magnetization rotation versus the field direction for rotating field magnitudes of $\mu_0 H =$ (*1*) 0.06, (*2*) 0.03, and (*3*) 0.006 T.

4. CONCLUSIONS

We have established that there are three remagnetization modes in a Fe/SmCo system grown on the (110) MgO substrate:

(1) a nonuniform small-scale remagnetization under strictly parallel orientation of the field, accompanied by the generation of spin springs of opposite chiralities;

(2) a quasi-uniform rotation of the net magnetization via a spin-spring mechanism when the magnetic field deviates from the easy axis;

(3) a uniform rotation of magnetization accompanied by the generation of a coherent spin spiral and a macroscopic domain structure.

We have shown that the tilt angle of the magnetic field with respect to the easy axis produces a very strong effect both on the magnitude of the remanent rotation angle of magnetization and on the choice of a right- or left-handed twisting of a spin spiral. The observed twisting asymmetry under a field reversal as well as under repeated remagnetization cycles gives evidence of the fact that the remagnetization processes in this system are largely determined by the moment of the initial deviation of the magnetization in the magnetically soft layer.

ACKNOWLEDGMENTS

We are grateful to S.D. Bader and J.S. Jiang for supplying the samples.

This work was supported in part by the program for basic research "New Materials and Structures" of the Division of Physics, Russian Academy of Sciences.

REFERENCES

1. B. Heinrich and J. F. Cochran, Adv. Phys. **42**, 523 (1999).
2. M. N. Baibich, J. M. Broto, A. Fert, *et al.*, Phys. Rev. Lett. **61**, 2472 (1988).
3. A. P. Malozemoff, J. Appl. Phys. **63**, 3874 (1988).
4. B. Dieny, J. P. Gavigan, and J. P. Rebouillat, J. Phys.: Condens. Matter **2**, 159 (1990).
5. S. S. P. Parkin, N. More, and K. P. Roche, Phys. Rev. Lett. **64**, 2304 (1990).
6. A. E. Berkowitz and K. Takano, J. Magn. Magn. Mater. **200**, 552 (1999).
7. J. Noguez and I. K. Schuller, J. Magn. Magn. Mater. **192**, 203 (1999).
8. E. Goto, N. Hayashi, T. Miyashita, *et al.*, J. Appl. Phys. **36**, 2951 (1965).
9. E. F. Kneller and R. Hawig, IEEE Trans. Magn. **27**, 3588 (1991).
10. D. Mauri, H.-C. Siegmann, P. S. Bagus, *et al.*, J. Appl. Phys. **62**, 3047 (1987).
11. E. E. Fullerton, J. S. Jiang, M. Grimsditch, *et al.*, Phys. Rev. B **58**, 12193 (1998).
12. A. Hubert, *Theorie der Domanenwande in Geordneten Medien* (Springer, Berlin, 1974; Mir, Moscow, 1977).
13. A. P. Malozemoff and J. C. Slonczewski, *Magnetic Domain Walls in Bubble Materials* (Academic, New York, 1979; Mir, Moscow, 1982).
14. V. S. Gornakov, L. M. Dedukh, V. I. Nikitenko, *et al.*, Zh. Éksp. Teor. Fiz. **86**, 1505 (1984) [Sov. Phys. JETP **59**, 881 (1984)].
15. V. S. Gornakov, V. I. Nikitenko, I. A. Prudnikov, *et al.*, Phys. Rev. B **46**, 10829 (1992).
16. J. S. Jiang, E. E. Fullerton, C. H. Sowers, *et al.*, IEEE Trans. Magn. **35**, 3229 (1999).
17. V. I. Nikitenko, V. S. Gornakov, Yu. P. Kabanov, *et al.*, J. Magn. Magn. Mater. **258–259**, 19 (2003).
18. V. K. Vlasko-Vlasov, U. Welp, J. S. Jiang, *et al.*, Phys. Rev. Lett. **86**, 4386 (2001).
19. J. C. Slonczewski, Phys. Rev. Lett. **67**, 3172 (1991).
20. L. A. Dorosinskii, M. V. Indenbom, V. I. Nikitenko, *et al.*, Physica C (Amsterdam) **203**, 149 (1992).
21. B. Heinrich, J. F. Cochran, M. Kowalewski, *et al.*, Phys. Rev. B **44**, 9348 (1991).
22. M. Ruhrig, R. Schafer, A. Hubert, *et al.*, Phys. Status Solidi A **125**, 635 (1991).
23. J. C. Slonczewski, J. Magn. Magn. Mater. **150**, 13 (1995).
24. L. M. Dedukh, V. N. Matveev, and V. I. Nikitenko, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 625 (2000) [JETP Lett. **71**, 430 (2000)].

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 99, No. 3, 2004, pp. 613–619.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 3, 2004, pp. 704–711.

SOLIDS
Electronic Properties

Fluctuations in the Paraphase of Improper Ferroelastics Taking into Account Indirect Interactions through the Field of Elastic Deformations

A. S. Yurkov
e-mail: fitec@omskcity.com
Received February 13, 2004

Abstract—The effect of indirect interactions (through the field of elastic deformations) on the temperature dependences of a two-point correlator of the order parameter of an improper ferroelastic is studied theoretically taking into account the interaction of fluctuations at different spatial points with one another and with defective elastic fields. The latter are accounted for by using a phenomenological field of the sources of defective elastic fields. Analysis is carried out using diagrammatic expansions followed by a transition to the Dyson equation. It is proposed that the Dyson equation be approximately solved nonperturbatively using the ansatz for an exact two-point Green function of the form $G_{\text{int}}(\mathbf{k}) = T/[\alpha_{ij}(\tau)k_i k_j + \beta(\tau)]$. Such an approach makes it possible to reduce the problem to solving a system of nonlinear algebraic equations, which can effectively be solved by numerical methods. The aggregate of the assumptions made is equivalent to the mean field theory, which, however, cannot be reduced in the present case to the Ginzburg–Landau theory in view of the essentially nonlocal character of the indirect interaction via the field of elastic deformations. The results of numerical calculations are considered for a defect-free Hg_2Cl_2 crystal, for which it is shown that parameters of dispersion α_{ij} acquire a substantial temperature dependence in a temperature range much broader than the width of the critical region of the given crystal. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is well known that fluctuations of the degrees of freedom corresponding to the order parameter sharply increase in systems experiencing a continuous phase transition (in particular, improper ferroelastics)[1] in the vicinity of T_c. It is assumed that the behavior of these fluctuations can be described in the Gaussian approximation (see, for example, [1]) in the framework of the Ginzburg–Landau theory if we disregard the narrow temperature interval directly adjoining T_c (critical region). In accordance with this theory, the spatial Fourier transform of the one-time correlator of fluctuations is a Lorentzian with an amplitude proportional to $(T - T_c)^{-1}$ and a width proportional to $(T - T_c)^{1/2}$.

At the same time, fluctuations of the order parameter in ferroelastics at $T > T_c$ were directly observed in experiments using diffuse scattering of X rays [2, 3]; in these experiments, the above-mentioned dependences have not generally been confirmed. It is important that this occurs in a wide temperature range clearly exceeding the critical region, which amount to only 1 K for Hg_2Cl_2 according to estimates from [4].

There are in principle at least two reasons for the discrepancy between the results of these experiment and the predictions of the Ginzburg–Landau theory. First, the contour of the line describing diffuse scattering of X rays cannot exactly correspond to the Fourier transform of the order parameter correlator (e.g., due to the effect of random elastic deformations that inevitably are present in ferroelastics). Second, the behavior of order parameter fluctuations may deviate from the predictions of the theory.

One of the factors due to which deviations in the behavior of fluctuations in improper ferroelastics may differ from the predictions of the Ginzburg–Landau theory can be the indirect interaction between fluctuations at different spatial points via the field of elastic deformations. Such an interaction is essentially nonlocal, which obviously contradicts the local character of the thermodynamic potential expansion normally used in the given theory.

It should be emphasized that nonlocal terms in the expansion of the crystal energy into a power series in the order parameter, which emerge as a result of the indirect interaction, are anharmonic. This complicates

[1] In a broad sense, the term improper ferroelastic can be applied to any crystal in which spontaneous deformation emerges as a result of a structural phase transition, the deformation itself being not an order parameter. Owing to striction coupling between polarization and deformation, ferroelectrics are also improper ferroelastics in the broad sense of the term. However, ferroelastics are usually the crystals in which electric polarization does not emerge in a structural phase transition. We will use here the term "improper ferroelastic" precisely in the narrow sense of the word since, in the case of ferroelectrics, we must consider, in addition to the interaction of fluctuations via the emerging deformation, the interaction via the macroscopic electric field appearing in ferroelectrics, which is beyond the scope of this article.

theoretical analysis and generally necessitates the use of certain approximations. Such approximations may be quite successful in the quantitative respect, especially if we use empirical and not calculated parameters; however, their role mainly lies in the description of the observed phenomena on a qualitative level.

Naturally, far away from the phase-transition temperature, when fluctuations are small, the interaction via the elastic field can be accounted for, as a rule, in perturbation theory. It is well known, however [5], that the effective parameter of the expansion in perturbation theory increases as $T \longrightarrow T_c$. Consequently, in the vicinity of the transition temperature, a finite order of perturbation theory does not provide even a qualitatively correct pattern. For this reason, it would be desirable to take into account (at least partly) the indirect interaction of fluctuations even in the zeroth approximation. This could extend the temperature range in which the theory gives physically reasonable dependences of the observables. In addition, elastic fields of crystal-structure defects of a real crystal may also play a significant role and should also be included in the description.

Thus, analysis of the role of elastic fields in the formation of the correlator of fluctuations in ferroelastics is quite important. It is also interesting to study the effect of elastic fields on the correlation between the line contour of diffuse X-ray scattering and the Fourier transform of one-time two-point correlator of fluctuations; however, this question is beyond the scope of this paper. We will only analyze the fluctuations of the order parameter themselves; the main attention will be paid to the qualitative aspect of the problem.

This article has the following structure. In Section 2, the adopted theoretical model is described taking into account the interaction between fluctuations located at different spatial points via an elastic field as well as the indirect elastic interaction of the fluctuating field of the order parameter with crystal structure defects. The latter are described phenomenologically by introducing a random field of the sources of elastic deformations. In this section, the diagrammatic technique that will be used in the subsequent section is introduced. In Section 3, we introduce the Dyson equation for the two-point correlator of fluctuations, which is solved by a nonperturbative method using the ansatz that has the following form in the **k** representation:

$$G_{\text{int}}(\mathbf{k}) = T/[\alpha_{ij}(\tau)k_i k_j + \beta(\tau)].$$

The approximation used in this case is equivalent to the mean field theory; however, in the present case this approximation cannot be reduced to the Ginzburg–Landau theory in view of the nonlocal character of the interaction. As a result, after numerical solution of the system of nonlinear algebraic equations obtained for a defect-free Hg_2Cl_2 crystal, it is shown that, in contrast to the Ginzburg–Landau theory, a strong temperature dependence of dispersion parameters α_{ij} emerges in addition to the temperature dependence of parameter β. The obtained results are discussed in Section 4.

2. EXPANSION OF THERMODYNAMIC POTENTIAL AND DIAGRAMMATIC TECHNIQUE

Since we are interested only in the qualitative aspect of the problem, we will confine our analysis to the case of a one-component order parameter η. It should be noted that the same approach could be used for crystals with a multicomponent order parameter if we assume that each component of η fluctuates independently, which seems to be a reasonable first approximation. A rigorous generalization of the theory to the case of a multicomponent order parameter can be carried out directly when necessary.

In the power expansion of the thermodynamic potential in spatially inhomogeneous order parameter η, we take into account harmonic (Gaussian) terms, fourth-order anharmonism in η, the interaction of the order parameter with the field of elastic deformation, and the interaction of elastic deformations with crystal-structure defects. We will describe the latter interaction by introducing the random field of sources of elastic stresses σ_{ij}. In the case of point defects, the field σ_{ij} is δ-correlated.

In accordance with the above arguments, we write the expansion of thermodynamic potential Φ in the form

$$\Phi = \Phi_2 + \Phi_{\text{el}} + \Phi_{\text{def}} + \Phi_4 + \Phi_{\text{int}}. \quad (1)$$

The terms in formula (1) are defined as

$$\Phi_2 = \frac{1}{2}\int(\lambda_{ij}\eta_{,i}\eta_{,j} + \tau\eta^2)d^3\mathbf{r}, \quad (2)$$

$$\Phi_{\text{el}} = \frac{1}{2}\int C_{ijkl}u_{i,j}u_{k,l}d^3\mathbf{r}, \quad (3)$$

$$\Phi_{\text{def}} = \int\sigma_{ij}u_{i,j}d^3\mathbf{r}, \quad (4)$$

$$\Phi_4 = \int g\eta^4 d^3\mathbf{r}, \quad (5)$$

$$\Phi_{\text{int}} = \int f_{ij}\eta^2 u_{i,j}d^3\mathbf{r}, \quad (6)$$

where the indices label spatial components; the index after coma indicates the partial derivative with respect to the corresponding spatial coordinate; recurrent indices indicate summation; η is the order parameter, u_i are elastic displacement components, C_{ijkl} is the elastic moduli tensor, λ_{ij} are the dispersion parameters, and g and f_{ij} are parameters of interactions.

Parameter τ requires additional explanation. In the Landau theory disregarding fluctuations, it is usually assumed that $\tau \propto (T - T_c)$. When the interaction between fluctuations is taken into account in the framework of the mean field theory, an additional correction to this term appears, which is proportional to T in the main approximation. In this case, we can assume that τ is a negative quantity independent of temperature; taking the fluctuation correction into account, we obtain $\tau \propto (T - T_c)$ as in the original phenomenological Landau theory.

It is important to note, however, that optical phonon modes of the crystal disregarded in potential (1) contribute to the temperature dependence of effective parameter τ in a completely analogous way. To take into account the effect of these modes phenomenologically, we assume that $\tau \propto (T - T_{c0})$, where T_{c0} is the "initial" phase transition temperature, which is higher than the actual phase-transition temperature of the crystal.

Obviously, taking into account the interactions, we can define the correlation function we are interested in via the continual integral,

$$G_{\text{int}}(\mathbf{r} - \mathbf{r}') = \frac{\int \eta(\mathbf{r})\eta(\mathbf{r}')e^{-\Phi/T}\mathcal{D}\eta\mathcal{D}u}{\int e^{-\Phi/T}\mathcal{D}\eta\mathcal{D}u}. \quad (7)$$

It should be emphasized that this relation defines the correlator for determinate field σ_{ij}. Actual physical quantities are defined by this correlator averaged over the set of realizations of the field σ_{ij}. It is important that averaging over σ_{ij} should be carried out precisely after evaluating the continual integral. If expression (7) did not contain the continual integral in the denominator, we could directly average expression (7). The presence of the denominator does not allow us to do this. One of the methods for averaging expressions of type (7) is the well-know replica technique (see, for example, [6]). However, in the theories based on diagrammatic expansions, the replica technique is equivalent to direct averaging of diagrams. It is well known that denominator in expressions of type (7) only leads to disappearance of disconnected diagrams and, hence, no serious difficulties necessitating the application of the replica technique emerge in such an approach.

Since the continual integral with respect to $\mathcal{D}u$ in Eq. (7) is Gaussian, continual integration with respect to $\mathcal{D}u$ can be carried out exactly, which leads to a formula similar to Eq. (7). The only difference is that this formula does not involve integration with respect to $\mathcal{D}u$ and contains, instead of potential (1), the effective potential

$$\Phi_{\text{eff}} = \Phi_2 + \Phi_4 + \Phi_u + \Phi_{\text{def-eff}} + \Phi_{\text{def-def}}, \quad (8)$$

where Φ_u describes indirect interaction between fluctuation via the field of elastic deformations, $\Phi_{\text{def-eff}}$ is a similar interaction of the order parameter with defects, and $\Phi_{\text{def-def}}$ is the elastic interaction between defects. The last term is insignificant for the subsequent analysis and will be omitted.

Indirect interaction are represented by the expressions

$$\Phi_u = \int \eta^2(\mathbf{r})U(\mathbf{r} - \mathbf{r}')\eta^2(\mathbf{r}')d^3\mathbf{r}d^3\mathbf{r}', \quad (9)$$

$$U(\mathbf{r}) = \frac{1}{2} f_{ij}R_{ijkl}(\mathbf{r})f_{kl}, \quad (10)$$

$$\Phi_{\text{def-eff}} = \int \eta^2(\mathbf{r})f_{ij}R_{ijkl}(\mathbf{r} - \mathbf{r}')\sigma_{kl}(\mathbf{r}')d^3\mathbf{r}d^3\mathbf{r}', \quad (11)$$

where $R_{ijkl}(\mathbf{r})$ is the elastic Green function defined by the formula

$$R_{ijkl}(\mathbf{r}) = -\frac{1}{(2\pi)^3}\int k_j k_l e^{i\mathbf{k}\cdot\mathbf{r}}\Pi_{ik}d^3\mathbf{k}, \quad (12)$$

$\Pi_{ij}(\mathbf{k})$ being the tensor reciprocal to tensor $\Gamma_{ij}(\mathbf{k}) = k_n k_m C_{injm}$.

Expanding in expression (7) the exponential of the terms describing the interaction into a series, we can easily construct the corresponding diagrammatic technique. In this technique, we depict the field σ_{ij} by a circle, the correlator (propagator) of the order parameter in the zeroth Gaussian approximation,

$$G(\mathbf{r} - \mathbf{r}') = \frac{1}{(2\pi)^3}\int \frac{T}{\lambda_{ij}k_i k_j + \tau}e^{i\mathbf{k}\cdot\mathbf{r}}d^3\mathbf{k}, \quad (13)$$

by a solid line, a nonlocal vertex of the indirect interaction $[-T^{-1}U(\mathbf{r})]$ by a wavy line, a nonlocal vertex of the indirect interaction with defects $[-T^{-1}f_{ij}R_{ijkl}(\mathbf{r} - \mathbf{r}')]$ by a spiral, and a local vertex $[-T^{-1}g]$ by a point.

Such a diagrammatic technique describes a crystal for a determinate field σ_{ij}. However, we require quantities averaged over realizations of σ_{ij}. This averaging slightly modifies diagrammatic rules, the modification affecting only the diagrams containing circles.

To determine the diagrammatic technique corresponding to the quantities averaged over field σ_{ij}, we consider Fig. 1, which contains several diagrams with elements affected by averaging over σ_{ij}. Additional wavy lines and fourth-order vertices do not change the following line of reasoning.

Each diagram in Fig. 1a corresponds to an analytic expression containing the number of factors σ_{ij} equal to the number of circles in the diagram. After averaging over realizations of fields of defects, these factors give correlators of field σ_{ij}. In the diagrammatic language, we will depict these correlators by circles corresponding to cofactors in the correlator enclosed in a rectan-

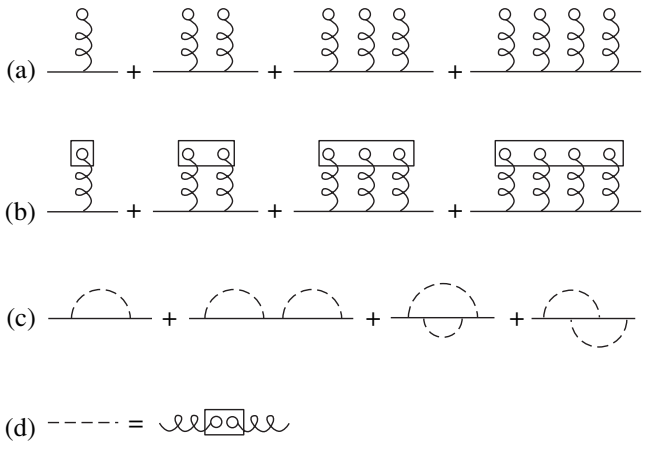

Fig. 1.

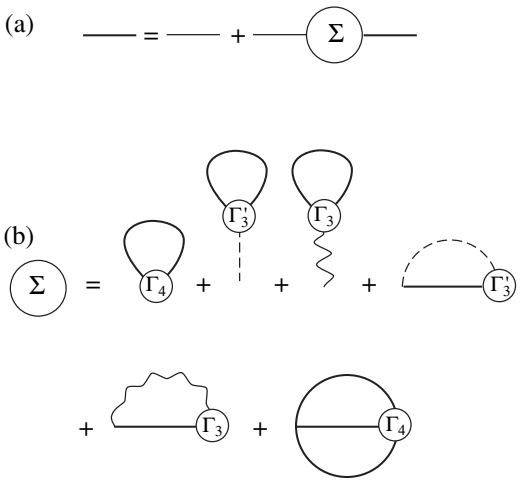

Fig. 2.

gle. As a result of averaging, the diagrams in Fig. 1a acquire the form shown in Fig. 1b.

Now, we must make a significant assumption according to which the field $\sigma_{ij}(\mathbf{r})$ is Gaussian with zero mean value (a nonzero mean would only renormalize the parameters, which is immaterial for our purposes). Then multipoint correlators of field σ_{ij} split into products of two-point correlators. As a result, the diagrammatic technique for quantities averaged over field σ_{ij} acquires an additional nonlocal vertex, which will be depicted by a broken line; circles and spirals disappear, and Fig. 1b is replaced by Fig. 1c. Figure 1d shows graphically how a vertex corresponding to a broken line is expressed in terms of the vertex of the indirect interaction with defects and the two-point correlator of field σ_{ij}. Analytically, this expression can be reduced to

$$V(\mathbf{r}-\mathbf{r}') = \int f_{ij} R_{ijkl}(\mathbf{r}-\mathbf{r}'') \frac{\langle \sigma_{kl}(\mathbf{r}'')\sigma_{nm}(\mathbf{r}''')\rangle}{T} \times R_{nmrs}(\mathbf{r}'''-\mathbf{r}') f_{rs} d^3\mathbf{r}'' d^3\mathbf{r}''', \quad (14)$$

where $T^{-1}V(\mathbf{r}-\mathbf{r}')$ is the analytic expression of a broken line.

3. DYSON EQUATION AND ITS APPROXIMATE NUMERICAL SOLUTION

Separating in the standard manner the class of one-particle irreducible diagrams (see, for example, [7]), introducing the exact propagator represented by a bold line, as well as the vertices renormalized by the interactions and depicted by circles with letter Γ, we can introduce the mass operator $\Sigma(\mathbf{k})$ and write the Dyson equation, which is shown graphically in Fig. 2. Exact vertices are now represented by infinite diagrammatic series, which, however, will not be written here.

If we could find exact vertices in closed form, the solution of the Dyson equation would lead to the exact solution of the problem since iterations of this equation generate the entire series of perturbation theories. However, the problem is that it is impossible even to write a closed equation for vertices. We can only write an infinite chain of equations in which lower order Green functions can be expressed in terms of higher order functions [8]. The Dyson equation given in Fig. 2 is just the first equation in this chain.

Thus, for vertices in the Dyson equation, we can either use several first terms of the corresponding series, or seek Γ by disentangling the above-mentioned chain of equations using some approximate method. In any case, we effectively sum not the entire series of perturbation theory, but only a certain infinite subsequence of the terms of this series.

Here, we will confine our analysis to the simplest approximation for vertices, taking them in the zeroth order in the interaction. In particular we disregard in this case the last diagram from those shown in Fig. 1c (this diagram contains a correction to the vertex), while the first three diagrams, as well as an infinitely large number of other higher order diagrams that do not contain corrections to vertices, are included.

For the mass operator $\Sigma(\mathbf{k})$, we also confine ourselves to the one-loop approximation; i.e., we disregard the last term in Fig. 2b. In addition, the first three terms in this figure give an identical expression to within constant coefficients. In fact, the action of the second and third diagrams boils down to a change in the value of g. We will henceforth assume that g is precisely the modified constant (which means that the second and third diagrams in Fig. 2b are implicitly taken into account).

In these approximations, the equations corresponding to the diagrams in Fig. 2 can be reduced to the relations

$$G_{\text{int}} = \frac{T}{\lambda_{ij}k_ik_j + \tau - T\Sigma(\mathbf{k})}, \tag{15}$$

$$\Sigma(\mathbf{k}) = -\frac{3g}{2\pi^3 T}\int G_{\text{int}}(\mathbf{q})d^3\mathbf{q} - \frac{1}{\pi^3 T}\int G_{\text{int}}(\mathbf{q}+\mathbf{k})[U(\mathbf{q}) - V(\mathbf{q})]d^3\mathbf{q}. \tag{16}$$

As regards the physical meaning of the approximations introduced above, we can make the following remarks. If the range of indirect interactions tends to zero, the second, third, fourth, and fifth terms on the right-hand side of Fig. 2b become absolutely equivalent to the first term. However, the inclusion of only this diagram in the zeroth approximation in corrections to the vertex leads to the conventional mean field theory equivalent to the Ginzburg–Landau theory.

Thus, we can state that the above approximations correspond to the mean field theory modified so that to take into account the nonlocal character of indirect interactions. Physically, this means that, in the case of a long-range interactions, the mean field acting on modes with different values of **k** also has different values depending on **k**.

To solve Eq. (15) approximately, we represent $G_{\text{int}}(\mathbf{k})$ in the form

$$G_{\text{int}}(\mathbf{k}) = \frac{T}{\alpha_{ij}(\tau)k_ik_j + \beta(\tau)}; \tag{17}$$

for the purpose of self-consistency, $\Sigma(\mathbf{k})$ should be represented in this case as a power expansion in components of vector **k** to within second-order terms. The terms linear in **k** turn out to be identically equal to zero since functions $U(\mathbf{q})$ and $V(\mathbf{q})$ are even. This leads to the following equations for functions $\alpha_{ij}(\tau)$ and $\beta(\tau)$:

$$\tau - T\Sigma(\mathbf{k} = 0) = \beta(\tau), \tag{18}$$

$$\lambda_{ij} - \frac{T}{2}\frac{\partial^2\Sigma}{\partial k_i\partial k_j}\bigg|_{\mathbf{k}=0} = \alpha_{ij}(\tau). \tag{19}$$

Direct calculation of $\Sigma(\mathbf{k} = 0)$ shows that the corresponding loop integrals linearly diverge at the upper limit (this reflects the well-known fact of infinite renormalization of the phase-transition temperature for the cutoff parameter tending to infinity). Consequently, the integrals should be truncated at the upper limit for a certain momentum equal to Λ.

In contrast to $\Sigma(\mathbf{k} = 0)$, quantity $\partial^2\Sigma/\partial k_i\partial k_j(\mathbf{k} = 0)$ can be expressed by an integral converging at the upper limit. This quantity is insensitive to truncation at large momenta, which is associated with the discrete structure of a real crystal. Consequently, calculating $\partial^2\Sigma/\partial k_i\partial k_j(\mathbf{k} = 0)$, we can assume that the maximal loop momentum is infinitely large. As a result of direct calculations, Eq. (19) is transformed to

$$\alpha_{ij} = \lambda_{ij} + \frac{T}{\pi^3}\int\frac{(4\alpha_{ri}\alpha_{sj} - \alpha_{rs}\alpha_{ij})q_rq_s - \alpha_{ij}\beta}{(\alpha_{rs}q_rq_s + \beta)^3} \times [U(\mathbf{q}) - V(\mathbf{q})]d^3\mathbf{q}. \tag{20}$$

Thus, the system of equations (18), (19) is uncoupled. We can seek coefficients α_{ij} as functions of β from Eq. (20) and only after that do we determine β as a function of temperature. Moreover, since β can be expressed in terms of an experimentally measured quantity (the amplitude of a diffuse reflection), we can obtain information on the temperature behavior of the fluctuation correlator even without solving Eq. (18).

It was mentioned above that $U(\mathbf{q})$ depends only on the direction of vector **q**. If the sources of random elastic field are δ-correlated, function $V(\mathbf{q})$ also possesses the same property. In this case, we can integrate Eq. (20) with respect to q in polar coordinates in the general form. To this end, we introduce the unit vector $\mathbf{p} = \mathbf{q}/q$; after a series of transformation, we obtain

$$\alpha_{ij} = \lambda_{ij} + \frac{T}{4\pi^2\beta^{1/2}}\int[U(\mathbf{p}) - V(\mathbf{p})] \times\frac{(3\alpha_{ri}\alpha_{sj} - \alpha_{ij}\alpha_{rs})p_rp_s}{(\alpha_{nm}p_np_m)^{5/2}}d\Omega_{\mathbf{p}}, \tag{21}$$

where $d\Omega_{\mathbf{p}}$ is the differential of the solid angle of directions of vector **p**.

System of equations (21) is nonlinear and can apparently be solved only numerically. In the numerical solution, however, it is more convenient to pass to dimensionless normalized quantities, writing α_{ij} in the form

$$\alpha_{\text{ij}} = \lambda_{ij} + A\int[U(\mathbf{p}) - C^{-1}V(\mathbf{p})] \times\frac{(3\alpha_{ri}\alpha_{sj} - \alpha_{ij}\alpha_{rs})p_rp_s}{(\alpha_{nm}p_np_m)^{5/2}}d\Omega_{\mathbf{p}}, \tag{22}$$

where

$$A = \frac{Tf^2}{4\pi^2\beta^{1/2}\lambda^{3/2}C}; \tag{23}$$

calculating $U(\mathbf{p})$ and $V(\mathbf{p})$, we must use elastic moduli C_{ijkl} normalized to the dimensional constant C and interaction parameters f_{ij} normalized to f. In this case, tensors α_{ij} and λ_{ij} in Eqs. (22) are normalized to dimensional constant λ. The temperature in the numerator on the right-hand side of Eq. (23) must be expressed in energy units.

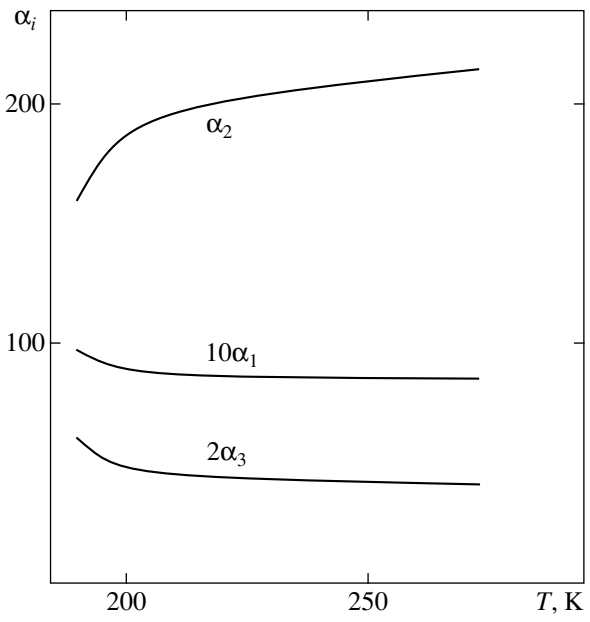

Fig. 3. Temperature dependence of dispersion parameters α_i of a Hg_2Cl_2 crystal. The curves for α_1 and α_3 are given on magnified scales (10 : 1 and 2 : 1, respectively) along the ordinate axis.

Let us obtain numerical estimates for a perfect (defect-free) Hg_2Cl_2 crystal. We will not try to fit the results to a specific experiment; the experimental results obtained in [2] will be used as a landmark for choosing approximations. Our aim is to estimate the effect of elastic interactions on the behavior of the correlator of fluctuations on the basis of experimentally obtained parameters, which are independent of experiments on X-ray scattering. Since the crystal defectiveness may change from sample to sample, it is meaningless to take defects into consideration.

First of all, we must verify that the temperature dependence of dispersion parameters α_{ij} predicted by the theory proposed here can be observed in a wide temperature range, which is considerably larger than the critical region. Consequently, we are interested in the temperature dependence of parameter β to a much smaller extent that the same dependence of parameters α_{ij}. In addition, it is well known [4, 9] that the fourth-order term in η renormalized by the elastic interaction are small for Hg_2Cl_2 crystals (at any rate, in the framework of the Landau phenomenological theory). For this reason, it seems expedient to set $\Sigma(\mathbf{k} = 0) = \text{const}$ in the first approximation and, hence, $\beta = \tau$ to within the renormalization of the phase-transition temperature. The temperature dependence of the amplitude of the diffuse X-ray reflection obtained in [2] also corresponds to this assumption, deviating from the proportionality to $1/(T - T_c)$ only in a small neighborhood of T_c.

Thus, the problem is actually reduced to solving the system of equations (22). The results of the numerical iterative solution of these equations using numerical values of parameters from [4, 9–11] are presented in Fig. 3. Tensor α_{ij} in this case is reduced to the principal axes rotated through 45° relative to tetragonal axes of the crystal, and only the strongest interaction with the ε_6 component of the strain tensor is taken into account [9].

In Fig. 3, parameters α_{ij} are normalized in such a way that their numerical values in the absence of the interaction with elastic deformations are 8, 255, and 18, which agrees with the results obtained in [10]. Pay attention to the fact that the scale on the ordinate axis, on which the temperature dependence of parameters α_1 and α_3 is depicted, is magnified tenfold and twofold, respectively, as compared to the scale for α_2. The curves are plotted for the temperature interval 190–273 K, which is beyond the critical region of the Hg_2Cl_2 crystal (according to the estimates obtained in [4], this regions lies between 0.1 and 1 K) in the vicinity of the phase-transition temperature $T_c = 186$ K.

It can be seen from the figure that parameter α_1 varies insignificantly. The variation of parameter α_2 is more pronounced (in the experimental temperature range, this parameter changes by a factor exceeding 1.3). Accordingly, the anomalous behavior of the width of the diffuse reflection must be more pronounced for scanning in the *ZE–X–EZ* direction than for scanning in the Γ–*X*–Γ direction. Parameter α_3 varies even more strongly (in the given temperature range, it changes by a factor of 1.5). Thus, the effects studied here must be manifested most clearly in diffuse X-ray scattering for scanning along the tetragonal axis of the crystal. Unfortunately, such scanning was not carried out in [2].

It should be emphasized once again that the above results were obtained taking into account the interaction with only one component ε_6 of the strain tensor, which is the strongest. In addition, the calculations were made only for a perfect (defect-free crystal) since it is difficult to obtain a priori estimates of the correlator for the sources of defective elastic fields. Such an approach is justified for estimating minimal possible effects; a meticulous comparison of the theory with experiment requires more detailed experimental information as compared to that provided in [2].

4. CONCLUSIONS

Thus, we have proved that the inclusion of the indirect interaction of order parameter fluctuations in improper ferroelastics via the field of elastic deformations radically changes (at least for a Hg_2Cl_2 crystal) the form of the temperature dependence of the Fourier transform of fluctuation correlator. Instead of temperature-independent coefficients of $\mathbf{k}^2$, the corresponding Lorentzian acquires the coefficients of $\mathbf{k}^2$ with a pronounced temperature dependence; this is observed in

the temperature interval in which the description in the framework of the mean field theory is applicable. This will naturally rule out the description of the behavior of diffuse X-ray scattering line width in terms of a simple power dependence. We can assume that either the power dependence obtained in [2] for the reflection width is an artifact of the method of experimental data processing or some processes that were disregarded here played a significant role in the experiment.

It should also be noted that the behavior of the amplitude of diffuse X-ray reflection in the proposed theory may completely correspond to the Ginzburg–Landau theory and can be described by a power dependence with an exponent of –1, in complete agreement with [2].

It should also be noted that neither the dynamic experiments on neutron scattering [10] nor the experiments on the Mandelstam–Brillouin scattering of light revealed any temperature dependence of the dispersion parameters in the corresponding Lorentzian. However, in such experiments, a high-frequency component of the temporal Fourier transform of the fluctuation correlator $\langle \eta(t, \mathbf{r})\eta(t', \mathbf{r}')\rangle$ is measured in actual practice. In other words, only rapid fluctuations play a noticeable role in such experiments. Considering that elastic deformations correspond to slow movements of the system, we cannot expect that the temperature behavior of the high-frequency part of fluctuations will be the same as the temperature dependence of the one-time correlator analyzed here. At the same time, diffuse X-ray scattering in which the frequency of a scattered photon is not detected is described precisely by the one-time correlator; it was shown above that the temperature dependence of the parameters of this correlator with allowance for elastic interaction differ from the predictions of the Ginzburg–Landau theory.

At the same time, final conclusions about the role of the processes considered here in real Hg_2Cl_2 crystals necessitates more detailed experiments and, above all, the analysis of the behavior of the diffuse X-ray reflection width under scanning along the tetragonal axis, where the most clear manifestation of the effects studied here can be expected. It would also be interesting to study experimentally the diffuse X-ray scattering on crystals with various levels of the defect concentration. We can expect that crystals containing defects will better exhibit effects similar to those calculated in the present study for a defect-free Hg_2Cl_2 crystal since the second term in the brackets in Eq. (22), which is proportional to the defect concentration, is quite analogous to the first term describing the elastic interaction in a defect-free crystal.

ACKNOWLEDGMENTS

The author thanks Yu.F. Markov for fruitful discussions.

REFERENCES

1. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics*, 3rd ed. (Nauka, Moscow, 1976; Pergamon Press, Oxford, 1980), Part 1.
2. Yu. F. Markov and K. Knorr, Fiz. Tverd. Tela (St. Petersburg) **41** (1), 148 (1999) [Phys. Solid State **41**, 132 (1999)].
3. S. R. Andrews, J. Phys. C: Solid State Phys. **19**, 3721 (1986); U. J. Nicholls and R. A. Cowley, J. Phys. C: Solid State Phys. **20**, 3417 (1987); A. Gibaud, R. A. Cowley, and P. W. Mitchell, J. Phys. C: Solid State Phys. **20**, 3849 (1987).
4. B. S. Zadokhin, Yu. F. Markov, and A. S. Yurkov, Zh. Éksp. Teor. Fiz. **104**, 2799 (1993) [JETP **77**, 286 (1993)].
5. V. G. Vaks, *Introduction to the Microscopic Theory of Ferroelectrics* (Nauka, Moscow, 1973) [in Russian].
6. V. S. Dotsenko, Usp. Fiz. Nauk **163** (6), 1 (1993) [Phys. Usp. **36**, 455 (1993)].
7. E. M. Lifshitz and L. P. Pitaevskiĭ, *Statistical Physics* (Fizmatlit, Moscow, 2002; Butterworth, Oxford, 1998), Part 2.
8. N. N. Bogolyubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields*, 4th ed. (Nauka, Moscow, 1984; Wiley, New York, 1980).
9. M. E. Boiko, Yu. F. Markov, V. S. Vikhnin, *et al.*, Ferroelectrics **130**, 263 (1992).
10. J. P. Benoit, G. Hauret, and J. Levebvre, J. Phys. (Paris) **43**, 641 (1982).
11. *Acoustic Crystals: Handbook*, Ed. by M. N. Shaskol'skaya (Nauka, Moscow, 1982) [in Russian].

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 99, No. 3, 2004, pp. 620–632.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 3, 2004, pp. 712–726.

SOLIDS
Electronic Properties

A Microscopic Model of Sequential Resonant Tunneling Transport through Weakly Coupled Superlattices

M. P. Telenkov and Yu. A. Mityagin

Lebedev Physical Institute, Russian Academy of Sciences, Moscow, 119991 Russia
e-mail: mityagin@mail1.lebedev.ru

Received December 28, 2003

Abstract—A microscopic model is developed for resonant tunneling transport in weakly coupled semiconductor superlattices in a constant external electric field. The model takes into account multiple subbands and electric-field dependence of scattering by acoustic and optical phonons, charged impurities, and interface roughness. The model is used as a basis for computing the resonant-tunneling profiles for structures with small size-quantization energies. The computed results are in good agreement with experiment. In structures of this type, an important role is played by electric-field dependence of scattering processes and the threshold behavior of elastic processes is strongly manifested. A substantial asymmetry is predicted not only for the first tunneling resonance, but also for higher order resonant tunneling processes. *© 2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The considerable current interest in transverse transport through semiconductor superlattices and quantum-well structures is primarily explained by the application of resonant tunneling through structures of this type in selective pumping of upper size-quantized subbands, inversion of subband populations, and development of unipolar injection infrared lasers based on intersubband optical transitions [1]. Furthermore, new phenomena associated with the resonant tunneling nature of current in such structures have been discovered, such as current multistability [2, 3] and self-sustained high-frequency current oscillations in a constant transverse electric field [4–6]. The latter phenomenon is of both fundamental and applied importance, because the oscillation frequency can vary within a very wide range.

The phenomena mentioned above are explained by the resonant tunneling nature of the current. In tunneling transport, an essential role is played by carrier-scattering processes, which determine both tunneling-current amplitude and resonant-tunneling profile. Therefore, mathematical modeling and understanding of the physics of the phenomena in question must rely on a sufficiently realistic microscopic model of resonant tunneling transport that takes into account various carrier-scattering processes.

A theory of this kind can be constructed by invoking mathematical methods based on Green's functions [7]. This approach was applied to the problem of tunneling transport through weakly coupled superlattices in a uniform constant electric field in [8–10], where one-electron two-time Green's functions were applied to model scattering processes without using any adjustable parameters and to obtain tunneling current densities that agree with experimental results.

In this paper, we use a similar approach to develop a microscopic model of resonant tunneling transport through superlattices and quantum-well structures. The model can be used to calculate tunneling current densities and resonant-tunneling profiles *ab initio* for a broad class of superlattices, including structures with small size-quantization energies (i.e., with wide quantum wells).

Wide quantum wells are characterized by multiple low-lying subbands separated by relatively small gaps. As a consequence, the coupling between different subband states induced by electric field in such wells leads to dependence of both resonant tunneling mechanisms and scattering processes on electric field. For this reason, we take into account the field dependence not only of tunneling mechanisms (as in [8–10]), but also of scattering processes in wells with multiple subbands. We analyze scattering by impurities, acoustic and optical phonons, and interface roughness.

The analysis of scattering mediated by charged impurities takes into account their screening. To calculate the screened potential of an array of impurities in a superlattice, we solve a three-dimensional problem by the Hartree self-consistent field method, taking into account multiple subbands. The mathematical apparatus developed in this study makes it possible to take into account the effect of electric field on scattering processes of this kind. It can also be used to allow for the electric-field effect on screening, which cannot be done by calculating matrix elements for zero field and then changing to a new basis. Moreover, we do not employ any approximation of the type usually employed (Tho-

mas–Fermi model). According to a numerical analysis, this is an important advantage to the present treatment, because we can therefore model the behavior of the polarization operator in the vicinity of the thresholds of elastic processes, above which a drastic decrease in screening is observed. As a result, we predict stronger scattering and a more complicated physical model. Furthermore, our model can be used to perform computations for arbitrary distributions of impurities both along the *z* axis and over the planar surface. Numerical results are obtained for structures with uniform bulk doping over the entire superlattice, instead of standard delta-doped structures.

Phonon-assisted scattering is modeled here approximately by applying the Fermi rule as done in the mathematical apparatus developed in [11]. We allow for the dependence of self-energy on the planar wave vector, which is important in certain cases, and on optical-phonon-assisted intrasubband scattering, which plays an important role in resonance "tail" effects.

Scattering by interface roughness is modeled here by developing the statistical approach of [8] to obtain analytical expressions for self-energy for various channels of this scattering mechanism in the Born approximation as functions of total energy, planar wave vector, and electric field. These expressions not only provide a basis for qualitative analysis, but also simplify numerical computations (in particular, they substantially reduce time complexity). We also propose a phenomenological approach that can be used to model correlations between different interfaces.

In addition, the present model takes into account the subband nonparabolicity (except for its role in scattering by interface roughness) and correlation between planar and transverse motions (dependence of the wavefunctions of transverse motion on the planar wave vector). This may be important for analyzing transport in cases when the subband bottom energy is close to the barrier height.

The theory developed here was used as a basis for computing resonant-tunneling profiles for structures with small size-quantization energies, and good agreement was achieved. The resulting widths of not only first-order, but also higher order, tunneling resonances are much greater than those estimated by using the uncertainty relation and relaxation times given by the Fermi rule [11–14]. Moreover, a substantial asymmetry is predicted for tunneling resonances of any order.

It was concluded in [13, 14] that optical-phonon-assisted intersubband scattering is much stronger for high-lying subbands, as compared to other scattering processes in structures with wide quantum wells. However, we show that comparable rates are characteristic of impurity-mediated intrasubband scattering for all subbands in such structures. Furthermore, it is found that the contributions of resonant tunneling transport to different subbands vary significantly, as manifested by dependence of characteristics of a tunneling resonance on its order.

We show that structures with small size-quantization energies are characterized by strong electric-field dependence of scattering processes and strong manifestations of the threshold behavior of elastic processes.

In Section 2, we present a general theory and consider scattering-assisted transport in uniform constant electric field without specifying any scattering mechanism. Sections 3, 4, and 5 are focused on scattering by impurities, optical and acoustic phonons, and interface roughness, respectively. Section 6 presents numerical results and their discussion.

2. BASIC ASSUMPTIONS AND MATHEMATICAL APPARATUS

The proposed model is based on the following assumptions. First, the time-independent and uniform external electric field **F** is applied in the negative direction of the growth axis of the structure (*z* axis). We also neglect all spin correlations in the electron gas; i.e., we consider a spin-degenerate system. Second, electron–electron interaction is taken into account only in calculating the screened potential of impurities by the Hartree self-consistent field method, so that analysis of superlattice transport is reduced to the problem of a noninteracting electron gas in an external potential. Third, tunneling coupling between electronic states in adjacent wells is treated as weak; i.e., the characteristic tunneling time is assumed to be longer than the intrawell relaxation time. This implies that the superelattice can be represented as an array of interacting quantum wells with strongly localized intrawell states. Therefore, equilibrium with the corresponding electrochemical potential is reached on the tunneling time scale in each well, and the tunneling probability per unit time is described by the Fermi rule for a system of noninteracting electrons with densities of states calculated by taking into account intrawell scattering. These densities of states are calculated by invoking two-time one-electron Green's functions [7, 15] without using any adjustable parameters. In these calculations, self-energies are obtained in the Born approximation, while the contribution of the interference effects due to various scattering processes to self-energy is neglected. Furthermore, we use the envelope function approximation [16] with a jump in effective mass across the well–barrier interface and neglect the effects due to the finite size of the superlattice, using periodic boundary conditions.

The original basis is defined as the product of the plane waves corresponding to transverse motion with the Wannier functions associated with motion along the growth axis:

$$\Psi^{\nu}_{n,\mathbf{k}}(\mathbf{r}, z) = \frac{\exp(i\mathbf{k}\cdot\mathbf{r})}{\sqrt{S}} W^{\nu}_{\mathbf{k}}(z - nd), \qquad (1)$$

where **r** is the coordinate in the plane perpendicular to the growth axis (planar surface), **k** is the planar wave vector, S is the area of the sample,

$$W_{\mathbf{k}}^{\nu}(z-nd) = \frac{1}{\sqrt{N}} \sum_{-\pi/d < q \le \pi/d} \exp(-iqnd)\varphi_{q\mathbf{k}}^{\nu}(z)$$

is the Wannier function quasi-localized in the nth well that corresponds to the state of the ν th subband with the wave vector **k**, N is the number of wells in the superlattice, d is the superlattice period, q is the Bloch wave vector, and $\varphi_{q\mathbf{k}}^{\nu}(z)$ is the corresponding Bloch function. The constant phase factors of the Bloch functions are taken so as to satisfy the Kohn theorem [17]. Then, each Wannier function is a real one symmetric or antisymmetric about the center of the corresponding well and is characterized by the highest degree of localization in the well among all possible linear combinations of Bloch functions.

It was shown in [18] that the external electric field gives rise not only to resonant tunneling, but also to coupling between states in different subbands of the same well. Following [8], we allow for the coupling by incorporating it into one-electron states as we change to the electric-field-independent basis [18]

$$\left\{ \begin{array}{l} \tilde{\Psi}_{n\mathbf{k}}^{\nu}(z, \mathbf{r}, U_d) = \dfrac{\exp(i\mathbf{k}\cdot\mathbf{r})}{\sqrt{S}} \tilde{W}_{\mathbf{k}}^{\nu}(z-nd, U_d), \\ \tilde{W}_{\mathbf{k}}^{\nu}(z-nd, U_d) = \displaystyle\sum_{\mu} U_{\mathbf{k}}^{\mu\nu}(U_d) W_{\mathbf{k}}^{\mu}(z-nd) \end{array} \right\}_{n,\mathbf{k}}^{\nu}, \quad (2)$$

where the elements of the unitary matrix $\{U_{\mathbf{k}}^{\nu\mu}(U_d)\}^{\nu\mu}$ satisfy the system of equations

$$\sum_{\nu'} \{[E^{\nu}(\mathbf{k}) - \tilde{E}^{\mu}(\mathbf{k})]\delta^{\nu\nu'} - U_d R_0^{\nu\nu'}\} U_{\mathbf{k}}^{\nu'\mu}(U_d) = 0. \quad (3)$$

Here,

$$E^{\nu}(\mathbf{k}) = \frac{1}{N} \sum_{-\pi/d < q \le \pi/d} E_{\mathbf{k}}^{\nu}(q)$$

denotes the energy $E_{\mathbf{k}}^{\nu}(q)$ averaged over the ν th Bloch band corresponding to the planar wave vector **k** (the mean energy of an electron in the corresponding Wannier state in the absence of scattering and electric field). The matrix

$$R_0^{\nu'\nu}(\mathbf{k}) = \int dz W_{\mathbf{k}}^{\nu'}(z) \frac{z}{d} W_{\mathbf{k}}^{\nu}(z) \quad (4)$$

represents the electric-field-induced coupling between subband states ν and ν' associated with **k** and localized in the same well, and $U_d = eFd$ is the voltage drop per superlattice period (e is the electron charge).

We not only renormalize the one-electron states in the absence of scattering and resonant tunneling matrix elements, as in [8–10], but also use basis (2) to calculate the self-energies for intrawell scattering processes (even in calculating the screened impurity potential). In other words, the dependence of scattering processes on the electric field is modeled almost exactly. A numerical analysis shows that this is particularly important for the modeling of transport in structures with small size-quantization energies and tunneling to high-lying subbands.

In the presence of external electric field and stationary defects characterized by the total scattering potential

$$V_{\mathrm{sc}}(\mathbf{r}, z) = \sum_i V_{\mathrm{sc},i}(\mathbf{r}, z),$$

where $V_{\mathrm{sc},i}(\mathbf{r}, z)$ is the total potential generated by defects of the ith type, the superlattice electron-gas Hamiltonian has the form

$$\tilde{H} = \tilde{H}_0 + \tilde{H}_{\mathrm{inter}}^{\mathrm{res}} + \tilde{H}_{\mathrm{sc}}, \quad (5)$$

where

$$\tilde{H}_0 = \sum_{n,\nu,\mathbf{k},\sigma} \{\tilde{E}_n^{\nu}(\mathbf{k})\} a_{n\nu\sigma}^{\dagger}(\mathbf{k}) a_{n\nu\sigma}(\mathbf{k}), \quad (6)$$

$$\tilde{H}_{\mathrm{inter}}^{\mathrm{res}} = \sum_{\substack{n,h>0,\\ \nu,\nu',\mathbf{k},\sigma}} \{\tilde{T}_h^{\nu'\nu}(\mathbf{k}) - U_d \tilde{R}_h^{\nu'\nu}(\mathbf{k})\} \times [a_{n+h,\nu'\sigma}^{\dagger}(\mathbf{k}) a_{n,\nu\sigma}(\mathbf{k}) + a_{n,\nu\sigma}^{\dagger}(\mathbf{k}) a_{n+h,\nu'\sigma}(\mathbf{k})] \quad (7)$$

is the part responsible for resonant tunneling (with conserved planar wave vector) induced by the superlattice potential and electric field, and

$$\tilde{H}_{\mathrm{sc}} = \sum_{\substack{i,n,h,\nu,\nu',\mathbf{k},\mathbf{p},\sigma \\ (h,\nu',\mathbf{p}) \neq (0,\nu,0)}} [\tilde{V}_{\mathrm{sc},i}]_{n+h,n}^{\nu',\nu}(\mathbf{k}+\mathbf{p}, \mathbf{k}) \times a_{(n+h)\nu'\sigma}^{\dagger}(\mathbf{k}+\mathbf{p}) a_{n\nu\sigma}(\mathbf{k}) \quad (8)$$

is the contribution corresponding to scattering by stationary superlattice defects. Here, $\alpha_{n\nu\sigma}^{\dagger}$ and $a_{n\nu\sigma}$ are the creation and annihilation operators for electrons in the

basis state characterized by a set quantum numbers $\{n, \nu, \sigma\}$ (σ corresponds to spin),

$$\tilde{E}_n^\nu(\mathbf{k}) = \tilde{E}^\nu(\mathbf{k}) - U_d n + \sum_i [\tilde{V}_{\mathrm{sc},i}]_{n,n}^{\nu,\nu}(\mathbf{k},\mathbf{k}), \quad (9)$$

$$\tilde{T}_h^{\nu'\nu}(\mathbf{k}) = \frac{1}{N}\sum_{\nu_1} U_{\mathbf{k}}^{\nu_1\nu'} \times \sum_{-\pi/d < q \le \pi/d} E_{\mathbf{k}}^{\nu_1}(q)\cos(qhd)U_{\mathbf{k}}^{\nu_1\nu}, \quad (10)$$

$$\tilde{R}_h^{\nu'\nu}(\mathbf{k}) = \int dz \tilde{W}_{\mathbf{k}}^{\nu'}(z-hd)\frac{z}{d}\tilde{W}_{\mathbf{k}}^{\nu}(z), \quad (11)$$

$$[\tilde{V}_{\mathrm{sc},i}]_{n+h,n}^{\nu',\nu}(\mathbf{k}+\mathbf{p},\mathbf{k}) = \int dz \tilde{W}_{\mathbf{k}+\mathbf{p}}^{\nu'}(z-hd) \times \frac{1}{S}\int d^2r \exp\{-i\mathbf{p}\cdot\mathbf{r}\}V_{\mathrm{sc},i}(z+nd,\mathbf{r})\tilde{W}_{\mathbf{k}}^{\nu}(z). \quad (12)$$

Under the assumptions made above, the total current density from nth to $(n+1)$th well is expressed as follows (see [19]):

$$j_{n\to n+1} = \frac{1}{S}\frac{4\pi e}{\hbar} + \sum_{\substack{\nu,\nu' \\ \mathbf{k},\mathbf{p}}} \left| \tilde{T}_1^{\nu'\nu}(\mathbf{k})\delta_{\mathbf{p},0} - U_d\tilde{R}_1^{\nu'\nu}(\mathbf{k})\delta_{\mathbf{p},0} + \sum_i [\tilde{V}_{\mathrm{sc},i}]_{n+1,n}^{\nu'\nu}(\mathbf{k}+\mathbf{p},\mathbf{k}) \right|^2 \times \rho_{\mathrm{eff}}(n,\nu,\mathbf{k},n+1,\nu',\mathbf{k}+\mathbf{p}), \quad (13)$$

where

$$\rho_{\mathrm{eff}}(n,\nu,\mathbf{k},n',\nu',\mathbf{k}') = \int_{-\infty}^{\infty} dE \rho_n^\nu(\mathbf{k},E)\rho_{n'}^{\nu'}(\mathbf{k}',E)[N_n(E) - N_{n'}(E)], \quad (14)$$

$$\rho_n^\nu(\mathbf{k},E) = \frac{1}{\pi} \times \frac{\mathrm{Im}\Sigma_n^\nu(\mathbf{k},E)}{[E - \tilde{E}_n^\nu(\mathbf{k}) - \mathrm{Re}\Sigma_n^\nu(\mathbf{k},E)]^2 + [\mathrm{Im}\Sigma_n^\nu(\mathbf{k},E)]^2} \quad (15)$$

is the contribution of the state $\{\nu, \mathbf{k}\}$ localized in the nth well to the corresponding density of states per spin projection (partial density of states per spin projection for the state $\{\nu, \mathbf{k}\}$ in the nth well),

$$\Sigma_n^\nu(\mathbf{k},E) = \sum_i \Sigma_{ni}^\nu(\mathbf{k},E), \quad (16)$$

$$\Sigma_{ni}^\nu(\mathbf{k},E) = \sum_{\substack{\nu_1,\mathbf{k}_1 \\ (\nu_1,\mathbf{k}_1)\neq(\nu,\mathbf{k})}} \frac{\left|[\tilde{V}_{\mathrm{sc},i}]_{n,n}^{\nu_1\nu}(\mathbf{k}_1,\mathbf{k})\right|^2}{E - \tilde{E}_n^{\nu_1}(\mathbf{k}_1) - i0} \quad (17)$$

is the contribution of scattering between states localized in the nth well to the self-energy of scattering processes, and $N_n(E)$ is the Fermi distribution for electrons in the nth well.

In what follows, we use the matrix elements for the potentials generated by stationary scattering centers and their squared absolute values averaged over their random distribution.

3. SCATTERING BY IONIZED IMPURITIES

The scattering impurities are treated as identical point charges Ze. Electron–electron interaction is taken into account only in calculating the screened potential of impurities by the Hartree self-consistent field method [15]. When the effects of tunneling and intrawell scattering due to other mechanisms on screening are neglected, the zeroth approximation used in the calculation of the screened impurity potential is described by the Hamiltonian

$$H^0 = \sum_{n,\mu,\mathbf{k},\sigma} [\tilde{E}_n^\nu(\mathbf{k})]a_{n\mu\sigma}^\dagger(\mathbf{k})a_{n\mu\sigma}(\mathbf{k}), \quad (18)$$

where $\tilde{E}_n^\nu(\mathbf{k}) = \tilde{E}^\nu(\mathbf{k}) - U_d n$ and the Fermi distribution over one-electron states in each well, $N_F(n, \mu, \mathbf{k})$. The field generated by impurities is instantly switched on at $t = 0$, and the screened impurity potential is sought at $t = \infty$.

Calculating the self-consistent response of the electron gas to an external disturbance (see [15]), we obtain a system of linear algebraic equations for the coefficients $V_{\mathrm{imp}}^{\mathrm{tot}}(n, \nu, \mathbf{k})$ of the decomposition of the screened impurity potential $V_{\mathrm{imp}}^{\mathrm{tot}}(\mathbf{r}, z)$ in basis (2):

$$\sum_{n',\nu'}\left\{-k^2\delta_{nn'}\delta^{\nu\nu'} + K_2^{\nu\nu'}(n-n',\mathbf{k}) + \frac{4\pi e^2}{\varepsilon}\Lambda_{nn'}^{\nu\nu'}(\mathbf{k})\right\}V_{\mathrm{imp}}^{\mathrm{tot}}(n',\nu',\mathbf{k}) = \frac{4\pi Ze^2}{\varepsilon\sqrt{S}}\sum_{i=1}^{N_{\mathrm{imp}}}\exp\{-i\mathbf{k}\cdot\mathbf{r}_i\}W_{\mathbf{k}}^\nu(z_i - nd). \quad (19)$$

The relation between the coefficients and the matrix element for the screened impurity potential follows directly from their definitions:

$$\langle n, \nu, \mathbf{k}|V_{\text{imp}}^{\text{tot}}|n', \nu', \mathbf{k}\rangle = \frac{1}{\sqrt{S}}\sum_{n_1, \nu_1} K_3^{\nu\nu'\nu_1}(n'-n, n_1-n, \mathbf{k}, \mathbf{k}', \mathbf{k}-\mathbf{k}') \times V_{\text{imp}}^{\text{tot}}(n_1, \nu_1, \mathbf{k}-\mathbf{k}'). \quad (20)$$

Here, ε is the superlattice permittivity, N_{imp} is the number of impurities, the coordinates of the ith impurity are $\mathbf{r} = \mathbf{r}_i$ and $z = z_i$,

$$K_2^{\nu\nu'}(n-n', \mathbf{k}) = \int dz \tilde{W}_{\mathbf{k}}^{\nu}(z-(n-n')d)\frac{\partial^2 \tilde{W}_{\mathbf{k}}^{\nu'}(z)}{\partial z^2}, \quad (21)$$

$$\Lambda_{nn'}^{\nu\nu'}(\mathbf{k}) = \sum_{\substack{n_1, \nu_1 \\ n_2, \nu_2}} \frac{2}{S}\sum_{\mathbf{p}} K_3^{\nu\nu_1\nu_2}(n_1-n, n_2-n, \mathbf{k}, \mathbf{k}+\mathbf{p}, \mathbf{p}) \times \frac{N_F(n_1, \nu_1, \mathbf{k}+\mathbf{p}) - N_F(n_2, \nu_2, \mathbf{p})}{\tilde{E}_{n_1}^{\nu_1}(\mathbf{k}+\mathbf{p}) - \tilde{E}_{n_2}^{\nu_2}(\mathbf{p}) - i0} \times K_3^{\nu_1\nu_2\nu'}(n_2-n_1, n'-n_1, \mathbf{k}+\mathbf{p}, \mathbf{p}, \mathbf{k}), \quad (22)$$

$$K_3^{\nu_1\nu_2\nu_3}(n_2-n_1, n_3-n_1, \mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \int dz \tilde{W}_{\mathbf{k}_1}^{\nu_1}(z)\tilde{W}_{\mathbf{k}_2}^{\nu_2}(z-(n_2-n_1)d) \times \tilde{W}_{\mathbf{k}_3}^{\nu_3}(z-(n_3-n_1)d). \quad (23)$$

In averaging over the impurities, they are assumed to be distributed independently and described by a uniform distribution function in the layer plane:

$$F(\{\mathbf{r}_i, z_i\}_i) = \prod_{i=1}^{N_{\text{imp}}} \frac{f(z_i)}{S}.$$

The sample area and impurity concentration are supposed to be such that the total probability of configurations in which almost all impurities are separated by distances greater than the effective screening radius is close to unity. This makes it possible to neglect the effects produced on the screening of the field of a particular impurity by the remaining impurities.

Under these assumptions, the diagonal matrix element for the screened potential of an array of impurities and squared absolute value of the matrix element for scattering by impurities between states in a particular well averaged over the impurities are expressed, respectively, as

$$\langle\langle n, \nu, \mathbf{k}|V_{\text{imp}}^{\text{tot}}|n, \nu, \mathbf{k}\rangle\rangle_{\text{imp}} = N_{\text{imp}}\int dz \langle n, \nu, \mathbf{k}|V_{\text{imp}}^{\text{tot}}(\mathbf{r}_i = 0, z_i = z)|n, \nu, \mathbf{k}\rangle f(z) \quad (24)$$

and

$$\left\langle \left|\langle n, \nu, \mathbf{k}|V_{\text{imp}}^{\text{tot}}|n, \nu_1, \mathbf{k}_1\rangle\right|^2\right\rangle_{\text{imp}} = N_{\text{imp}}\int dz \left|\langle n, \nu, \mathbf{k}|V_{\text{imp}}^{\text{tot}}(\mathbf{r}_i = 0, z_i = z)|n, \nu_1, \mathbf{k}_1\rangle\right|^2 f(z). \quad (25)$$

Here, $\langle n, \nu, \mathbf{k}|V_{\text{imp}}^{\text{tot}}(\mathbf{r}_i = 0, z_i = z)|n, \nu_1, \mathbf{k}_1\rangle$ is the scattering matrix element for the impurity located at $\mathbf{r}_i = 0$, $z_i = z$ for which the screening is calculated without taking into account the effects of the remaining impurities.

The mathematical apparatus developed here was used to allow for electric-field effects on scattering by impurities, which was not done in [8, 9]. A numerical analysis has shown that this leads to particularly important modifications in models of resonant-tunneling profiles for structures with small size-quantization energies and tunneling to high-lying subbands.

Note also that the effect of electric field on screening cannot be described by calculating the matrix element for a screened impurity potential for zero field (in basis (1)) and then renormalizing it by unitary transformation to basis (2).

Moreover, even though the thresholds of elastic processes manifest themselves as quasi-two-dimensional characteristics, the screening of the impurity potential is strongly affected by the possibility of motion along the growth axis. The most important manifestation of this effect is the strong dependence of polarization operator (22) on the wavefunctions corresponding to transverse motion, which vary substantially with the subband number. For structures with small size-quantization energies and tunneling to high-lying subbands, they also depend on the electric-field strength. This justifies the analysis of an essentially three-dimensional problem developed here (without any reference to quasi-two-dimensionality) to calculate the screened impurity potential for structures with small size-quantization energies.

In this study, we performed computations for a structure with uniform bulk doping (required for comparison with experiment). However, the model proposed here can be applied to arbitrary impurity distributions both along the z axis and over the planar surface, instead of standard delta-doped structures.

Furthermore, the present mathematical apparatus makes it possible to calculate the matrix element for a screened impurity potential (required for numerical analysis) directly, without calculating either the screened

potential in the coordinate representation or the matrix elements for the bare impurity potential.

4. PHONON-ASSISTED SCATTERING

In this study, we allow for scattering processes involving emission of longitudinal polar optical and long-wavelength acoustic phonons. Their contribution to the self-energy for intrawell scattering is modeled here approximately by applying the Fermi rule: the total self-energy for scattering by impurities and interface roughness from the ν*th* subband of the *n*th well with planar wave vector **k** is corrected by adding the independent of *E* term

$$\Sigma_{\mathrm{opt(ac)}}(n, \nu, \mathbf{k}) = \frac{i\hbar}{2\tau_{\mathrm{opt(ac)}}(n, \nu, \mathbf{k})}. \quad (26)$$

Here, the contribution $\tau^{-1}_{\mathrm{opt(ac)}}(n, \nu, \mathbf{k})$ of intrawell scattering involving emission of longitudinal polar optical phonons and long-wavelength acoustic phonons to the inverse lifetime for a particular state is calculated by using the model developed in [11], but with states (2) treated as one-electron ones.

In contrast to [8], we allow for electric-field dependence of these scattering mechanisms. This strongly modifies resonant-tunneling profiles for structures with small size-quantization energies. Moreover, we take into account phonon-assisted intrasubband scattering, which plays an important role in resonance tail effects, and the planar-wave-vector dependence of the contribution of these scattering processes to self-energy. These effects manifest themselves in numerical results.

5. SCATTERING BY INTERFACE ROUGHNESS

The well–barrier interface localized at a point $z = z_m$ on the *m*th interface of an ideal superlattice is represented by a fluctuating function $\xi_m(\mathbf{r})$. Interface fluctuations are described by a joint probability functional $P[\{\xi_m(\mathbf{r})\}_m]$, which is used to average all dependent functionals:

$$\langle F[\{\xi_m(\mathbf{r})\}_m]\rangle_{\mathrm{rough}} = \sum_{\{\xi_m(\mathbf{r})\}_m} P[\{\xi_m(\mathbf{r})\}_m]F[\{\xi_m(\mathbf{r})\}_m] \quad (27)$$

is the functional $F[\{\xi_m(\mathbf{r})\}_m]$ averaged over interface fluctuations.

From here on, scattering by interface roughness is modeled by using a parabolic-subband approximation [16]. Accordingly, the wavefunctions in original basis (1) corresponding to transverse motion are independent of the planar wave vector and are calculated for $\mathbf{k} = 0$. In the absence of scattering and electric field, the one-electron energy is a quadratic function of the planar wave vector:

$$E^{\nu}_n(\mathbf{k}) = E^{\nu}_n(\mathbf{k} = 0) + \frac{\hbar^2}{2N}\sum_q \left\langle \varphi^{\nu}_{q,0}\left|\frac{1}{\tilde{m}(z)}\right|\varphi^{\nu}_{q,0}\right\rangle \mathbf{k}^2, \quad (28)$$

where $\tilde{m}(z) = \tilde{m}_{\mathrm{w}}$ or $\tilde{m}(z) = \tilde{m}_{\mathrm{b}}$ if *z* belongs to the well or barrier, respectively. In addition to this approximation, we neglect the distinction of the integrals $\langle \varphi^{\nu}_{q,0}|1/\tilde{m}(z)|\varphi^{\nu}_{q,0}\rangle$ from $1/\tilde{m}_{\mathrm{w}}$ since the superlattice is weakly coupled. Then, the elements of the renormalizing matrix are independent of the planar wave vector, and so are the wavefunctions in basis (2), while the dispersion of one-electron energy with respect to the planar wave vector in the absence of scattering has the form

$$\tilde{E}^{\nu}_n(\mathbf{k}) = \tilde{E}^{\nu}_n(\mathbf{k} = 0) + E_{\mathbf{k}} \quad (29)$$

with $E_{\mathbf{k}} = \hbar^2\mathbf{k}^2/2\tilde{m}_{\mathrm{w}}$.

In these approximations, the contribution of impurities to the average value of one-electron energy (29) is also independent of the planar wave vector (which is included into $\tilde{E}^{\nu}_n(\mathbf{k} = 0)$ in what follows).

The potential due to interface roughness added to the superlattice potential is defined as follows (see [8]):

$$V^{\mathrm{rough}}(\mathbf{r}, z) = U_{\mathrm{SL}}\sum_m \xi_m(\mathbf{r})\delta(z - z_m), \quad (30)$$

where U_{SL} is the superlattice barrier height.

Then, after averaging over interface fluctuations, the matrix element corresponding to (30) and its squared absolute value are expressed, respectively, as

$$[V^{\mathrm{rough}}]^{\nu'\nu}_{n+h,n}(\mathbf{k} + \mathbf{p}, \mathbf{k}) = U_{\mathrm{SL}}\sum_m \tilde{W}^{\nu'}(z_m - nd - hd)\tilde{W}^{\nu}(z_m - nd) \times \langle\langle \exp(-i\mathbf{p}\cdot\mathbf{r})\xi_m(\mathbf{r})\rangle_{\mathbf{r}}\rangle_{\mathrm{rough}} \quad (31)$$

and

$$\left|[V^{\mathrm{rough}}]^{\nu'\nu}_{n+h,n}(\mathbf{k} + \mathbf{p}, \mathbf{k})\right|^2 = \sum_{m,m'} J^{\nu'\nu}_{n,h}(\mathbf{p}, m, m'), \quad (32)$$

where

$$J^{\nu'\nu}_{n,h}(\mathbf{p}, m, m') = U^2_{\mathrm{SL}}\tilde{W}^{\nu'}(z_m - (n + h)d) \times \tilde{W}^{\nu'}(z_{m'} - (n + h)d)\tilde{W}^{\nu}(z_m - nd)\tilde{W}^{\nu}(z_{m'} - nd) \times \langle \exp(i\mathbf{p}\cdot\mathbf{R})\langle\langle \xi_m(\mathbf{r})\xi_{m'}(\mathbf{r} + \mathbf{R})\rangle_{\mathbf{r}}\rangle_{\mathrm{rough}}\rangle_{\mathbf{R}} \quad (33)$$

and

$$\langle f(\mathbf{r})\rangle_{\mathbf{r}} = \frac{1}{S}\int d^2r f(\mathbf{r}).$$

Contributions (33) to squared absolute value (32) of the scattering matrix element with $m = m'$ and $m \neq m'$ have a different nature. The former quantities (to which the analyses in [8, 10] were restricted) represent direct scattering by particular interfaces and are determined by the corresponding autocorrelation functions. The latter contributions are associated with the interference of scattering by different interfaces and depend on the corresponding cross-correlation functions.

We assume that the autocorrelation function for each interface has the form (see [20])

$$\langle\langle \xi_m(\mathbf{r})\xi_m(\mathbf{r}+\mathbf{R})\rangle_{\mathbf{r}}\rangle_{\text{rough}} = \eta_m^2 \exp(-R/\lambda_m). \quad (34)$$

Then, the contribution of direct scattering by the mth interface to the squared absolute value of a matrix element averaged over interface fluctuations is expressed as follows [8]:

$$J^{\nu'\nu}_{n,h}((\lambda_m p)^2, m, m) = U_{\text{SL}}^2 [\tilde{W}^{\nu'}(z_m - (n+h)d)]^2 [\tilde{W}^{\nu}(z_m - nd)]^2 \times \eta_m^2 \frac{2\pi\lambda_m^2}{S} \frac{1}{[1+(\lambda_m p)^2]^{3/2}}. \quad (35)$$

For each interface, we suppose that

$$\langle\langle \xi_m(\mathbf{r})\rangle_{\mathbf{r}}\rangle_{\text{rough}} = 0, \quad (36)$$

which implies zero value of the diagonal matrix element in (31).

The resulting total contribution of scattering by interface roughness to self-energy calculated in the Born approximation is the sum of the contributions due to direct scattering by each particular interface (with $m = m'$) and the contributions of interference between scattering by different interfaces (with $m \neq m'$):

$$\Sigma^{\text{rough}}(n, \nu, \mathbf{k}, E) = \sum_{\substack{m, m' \\ \nu_1}} \Sigma^{\text{rough}, \nu_1}_{mm'}(n, \nu, \mathbf{k}, E). \quad (37)$$

In the analysis of intrawell scattering by interface roughness presented below, we neglect the real part of self-energy.

We obtained the following analytical expressions for the imaginary parts of direct contributions to the self-energy of scattering by interface roughness:

$$\text{Im}\Sigma^{\text{rough}, \nu_1}_{mm}(n, \nu, \mathbf{k}, E) = \frac{U_{\text{SL}}^2}{4E_{\lambda_m}} \eta_m^2 [\tilde{W}^{\nu}(z_m - nd)]^2 [\tilde{W}^{\nu_1}(z_m - nd)]^2 \times I_n^{\nu_1}(\lambda_m k, E, m, m)\theta(E - \tilde{E}_n^{\nu_1}(0)), \quad (38)$$

where

$$I_n^{\nu}(\lambda_m k, E, m, m) = 4[1 + (\lambda_m k + \sqrt{\kappa_n^{\nu}(E)})^2]^{-1} \times [1 + (\lambda_m k - \sqrt{\kappa_n^{\nu}(E)})^2]^{-1/2} \times E\left(\left[\frac{4\lambda_m k\sqrt{\kappa_n^{\nu}(E)}}{1 + (\lambda_m k + \sqrt{\kappa_n^{\nu}(E)})^2}\right]^{1/2}\right), \quad (39)$$

$\theta(x)$ is the Heaviside step function, $E(x)$ is the complete elliptic integral of the second kind, $\kappa_n^{\nu}(E) = [E - \tilde{E}_n^{\nu}(0)]/E_{\lambda_m}$, and $E_{\lambda_m} = \hbar^2/2m_w\lambda_m^2$ is the size-quantization energy for a system of characteristic size λ_m.

For arbitrary values of parameters, integral (39) can be represented by either the series

$$I_n^{\nu}(\lambda_m k, E, m, m) = 2\pi\left[1 + \left(\lambda_m k + \sqrt{\kappa_n^{\nu}(E)}\right)^2\right]^{-1}\left[1 + \left(\lambda_m k - \sqrt{\kappa_n^{\nu}(E)}\right)^2\right]^{-1/2} \times \left\{1 - \sum_{l=1}^{\infty} \frac{\prod_{i=0}^{l-1}(2i+1)^2}{(2l-1)\prod_{i=1}^{l}(2i)^2}\left[\frac{4\lambda_m k\sqrt{\kappa_n^{\nu}(E)}}{1 + (\lambda_m k + \sqrt{\kappa_n^{\nu}(E)})^2}\right]^l\right\} \quad (40)$$

or

$$I_n^{\nu}(\lambda_m k, E, m, m) = 4\pi\left[1 + \left(\lambda_m k + \sqrt{\kappa_n^{\nu}(E)}\right)^2\right]^{-1/2}\left[1 + \left(\lambda_m k - \sqrt{\kappa_n^{\nu}(E)}\right)^2\right]^{-1/2} \times \left\{\left[1 + \left(\lambda_m k + \sqrt{\kappa_n^{\nu}(E)}\right)^2\right]^{1/2} + \left[1 + \left(\lambda_m k - \sqrt{\kappa_n^{\nu}(E)}\right)^2\right]^{1/2}\right\}^{-1} \quad (41)$$

$$\times\left\{1+\frac{1}{4}[y_n^{\nu}(\lambda_m k, E, m)]^2+\sum_{l=2}^{\infty}\frac{\prod_{i=0}^{l-1}(2i+1)^2}{\prod_{i=0}^{l}(2i)^2}[y_n^{\nu}(\lambda_m k, E, m)]^{2l}\right\},$$

where

$$y_n^{\nu}(\lambda_m k, E, m) = \frac{2\left[1+\left(\lambda_m k+\sqrt{\kappa_n^{\nu}(E)}\right)^2\right]^{1/2}}{\left[1+\left(\lambda_m k+\sqrt{\kappa_n^{\nu}(E)}\right)^2\right]^{1/2}+\left[1+\left(\lambda_m k-\sqrt{\kappa_n^{\nu}(E)}\right)^2\right]^{1/2}}. \quad (42)$$

Both series expansions are exact ones with monotonically decreasing terms. Since the convergence rate of series (41) is higher than that of (40) for any particular set of parameter values, series (41) is better suited for numerical analysis, whereas (40) is better suited for analytical calculations.

Expressions (38)–(42) can be used to substantially simplify both algorithmic implementation and time complexity without any loss of accuracy. This made it possible to allow for the effects of electric field on scattering processes (which was not done in [8, 10]).

"Exact" modeling of interference of scattering by different interfaces is a more complicated task: while the autocorrelation functions can be evaluated experimentally (see [20]), no technique is currently available for evaluating the cross-correlation functions corresponding to different interfaces. However, there are reasons to believe that, at least, adjacent interfaces in real heterostructures are correlated to some extent. The phenomenological approach proposed below makes use of several adjustable parameters to take into account the interference effects due to scattering by different interfaces.

Since we deal with a weakly coupled superlattice, only the contributions of interference between adjacent well boundaries (labeled 1 and 2) and adjacent barrier boundaries (labeled 2 and 3) are taken into account in our models of intrawell and interwell scattering, respectively (see Fig. 1). We assume that the autocorrelation functions corresponding to interfaces 1, 2, and 3 are identical,

$$\langle\langle\xi_1(\mathbf{r})\xi_1(\mathbf{r}+\mathbf{R})\rangle_{\mathbf{r}}\rangle_{\text{rough}} = \langle\langle\xi_2(\mathbf{r})\xi_2(\mathbf{r}+\mathbf{R})\rangle_{\mathbf{r}}\rangle_{\text{rough}} = \langle\langle\xi_3(\mathbf{r})\xi_3(\mathbf{r}+\mathbf{R})\rangle_{\mathbf{r}}\rangle_{\text{rough}} \equiv A(\mathbf{R}), \quad (43)$$

and so are the cross-correlation functions,

$$\langle\langle\xi_1(\mathbf{r})\xi_2(\mathbf{r}+\mathbf{R})\rangle_{\mathbf{r}}\rangle_{\text{rough}} = \langle\langle\xi_2(\mathbf{r})\xi_3(\mathbf{r}+\mathbf{R})\rangle_{\mathbf{r}}\rangle_{\text{rough}} \equiv K(\mathbf{R}). \quad (44)$$

To characterize the relative contribution of interference between different interfaces, we introduce the normalized correlation function

$$g(\mathbf{p}) = \frac{\langle\exp\{i\mathbf{p}\cdot\mathbf{R}\}K(\mathbf{R})\rangle_{\mathbf{R}}}{\langle\exp\{i\mathbf{p}\cdot\mathbf{R}\}A(\mathbf{R})\rangle_{\mathbf{R}}}. \quad (45)$$

It is obvious that

$$|g(\mathbf{p})| < 1. \quad (46)$$

If the interfaces are totally uncorrelated, then $g(\mathbf{p}) = 0$.

If the interfaces are fully correlated (identical up to inversion), then $|g(\mathbf{p})| = 1$, we have the highest possible contribution of interference, and the corresponding term in the expression for the squared absolute value of the matrix element is

$$J_{n,n+1}^{\nu'\nu}((\lambda p)^2, m, m') = U_{\text{SL}}^2\tilde{W}^{\nu'}(z_m-d)\tilde{W}^{\nu'}(z_{m'}-d)\tilde{W}^{\nu}(z_m) \times\tilde{W}^{\nu}(z_{m'})\eta^2\frac{2\pi\lambda^2}{S}[1+(\lambda p)^2]^{-3/2}, \quad (47)$$

while the corresponding contribution to the imaginary

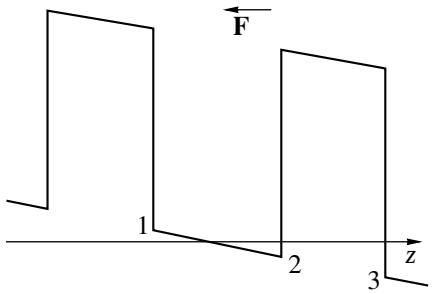

Fig. 1. Schematic of a superlattice in electric field.

Table 1

Order of resonance	j^{res}, mA/cm^2		Γ, meV		Γ_-/Γ_+	
	$n_{imp} = 1.75 \times 10^{10}$ cm^{-2}	$n_{imp} = 5 \times 10^{10}$ cm^{-2}	$n_{imp} = 1.75 \times 10^{10}$ cm^{-2}	$n_{imp} = 5 \times 10^{10}$ cm^{-2}	$n_{imp} = 1.75 \times 10^{10}$ cm^{-2}	$n_{imp} = 5 \times 10^{10}$ cm^{-2}
1	2.64	4.26	3.40	5.00	0.55	0.67
2	28.70	53.87	4.43	7.35	0.48	0.62
3	146.17	232.75	5.20	11.40	0.60	0.56
4	469.67	1045.59	9.05	11.15	1.03	0.86

part of self-energy is

$$\mathrm{Im}\Sigma_{mm'}^{\mathrm{rough},\,\nu_1}(n, \nu, \mathbf{k}, E) = \frac{U_{\mathrm{SL}}^2}{4E_\lambda}\eta^2\tilde{W}^{\nu}(z_m)$$

$$\times \tilde{W}^{\nu}(z_{m'})\tilde{W}^{\nu_1}(z_m)\tilde{W}^{\nu_1}(z_{m'}) \qquad (48)$$

$$\times I_n^{\nu_1}(\lambda k, E, m, m')\theta(E - \tilde{E}_n^{\nu_1}(0)).$$

The contributions due to direct scattering are modeled by using isotropic autocorrelation function (34); i.e., the interfaces are assumed to be isotropic on average. Extending this assumption to interference of different interfaces, we also use an isotropic cross-correlation function. Since it is a real-valued analytic function, the normalized correlation function depends on p^2 (its series expansion in terms of **p** contains only even powers of the absolute value). The wave-vector range that contributes most to scattering by interface roughness is determined by the correlation length: $\Delta p \propto \lambda^{-1}$. For correlation lengths varying between 10 and 100 Å and an electron concentration of about 10^{10} cm^{-2}, the inverse correlation length is $\lambda^{-1} \sim (0.01–0.1)k_F$, where k_F is the Fermi wave vector.

The Fourier transform of autocorrelation function (34),

$$\langle \exp\{i\mathbf{p}\cdot\mathbf{R}\}A(\mathbf{R})\rangle_{\mathbf{R}} = \eta^2\frac{2\pi\lambda^2}{S}[1 + (\lambda p)^2]^{-3/2}, \qquad (49)$$

weakly varies with λp. This allows us to assume that the Fourier transform of the cross-correlation function also is a weakly varying function of λp.

Accordingly, normalized correlation function (45) can be represented by a finite number of terms of its series expansion in terms of $(\lambda p)^2$,

$$g(p^2) \approx g_0 + g_1(\lambda p)^2 + g_2(\lambda p)^4 + \ldots, \qquad (50)$$

and the expansion coefficients can be treated as adjustable parameters.

As a result, we determine the approximate Fourier transform of the cross-correlation function directly from (45) and use it to calculate the corresponding contributions to (32) and (37).

According to (33), the role played by scattering by interface roughness strongly depends on barrier height and quantum-well width (which determine the values of wavefunctions on interfaces). Since the structures numerically analyzed in this study are characterized by a relatively small barrier height and large quantum-well width, scattering by interface roughness is relatively weak as compared to optical-phonon- and impurity-assisted scattering (being slightly stronger than acoustic-phonon-assisted scattering). With increasing barrier height or decreasing well width, the role of scattering by interface roughness increases and becomes important for a broad class of structures (see [21], where scattering times were given by the Fermi rule, mobility was calculated by using the Boltzmann equation, and Bloch states were taken as a basis). Therefore, the role played by interference between adjacent interfaces is an important issue and requires further analysis.

6. RESULTS AND DISCUSSION

Even though the model developed in this study can be used in computations for electric fields corresponding to resonances of all orders in the entire range of geometric parameters of weakly coupled structures, this paper is focused on structures with small size-quantization energies, for which the present model predicts a number of striking and unexpected phenomena. Figure 2 shows the tunneling current density calculated as a function of electric-field intensity by taking into account intrawell scattering assisted by longitudinal polar optical and long-wavelength acoustic phonons, impurities, and interface roughness for a weakly coupled $GaAs/Al_{0.3}Ga_{0.7}As$ superlattice (with a barrier width of 100 Å and a quantum-well width of 250 Å) with uniform bulk doping characterized by the impurity concentrations $n_{imp} = 1.75 \times 10^{10}$ and 5×10^{10} cm^{-2} at a temperature of 4.2 K. The interface roughness was parameterized by a correlation length of 70 Å and a mean amplitude of 3 Å, and we set $g(\mathbf{p}) = 0.5$. Table 1

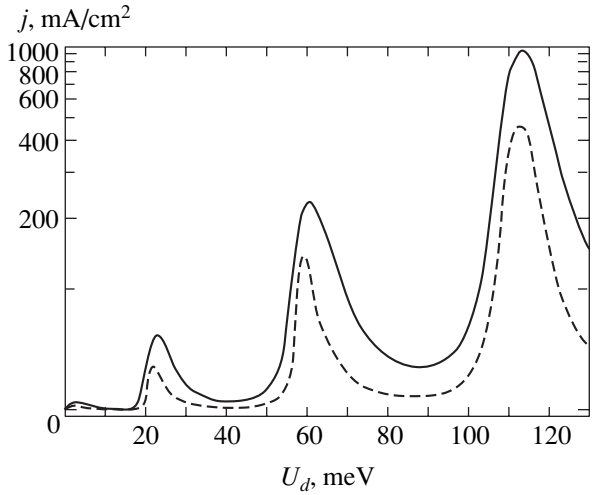

Fig. 2. Tunneling current density between adjacent quantum wells versus voltage drop per period for a $GaAs/Al_{0.3}Ga_{0.7}As$ superlattice with a well width of 250 Å and a barrier width of 100 Å: solid and dashed curves correspond to $n_{imp} = 5 \times 10^{10}$ and 1.75×10^{10} cm^{-2}, respectively.

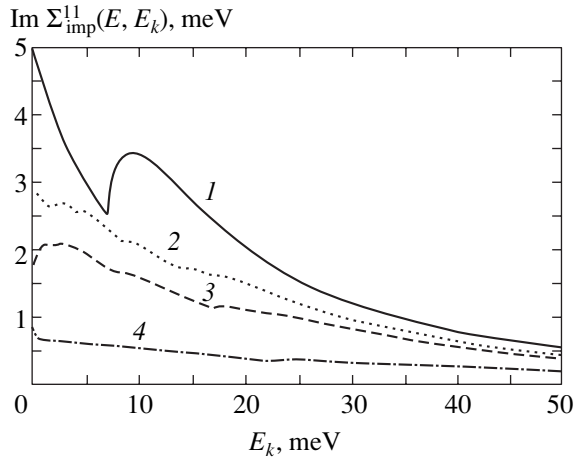

Fig. 3. Imaginary part of the contribution of impurity-assisted intrasubband scattering in the first subband to self-energy versus the planar energy $E_{\mathbf{k}} = \hbar^2\mathbf{k}^2/2m_w$ for the total energies $E = 0$ (*1*), 2 (*2*), 4 (*3*), and 10 meV (*4*). The total energies are measured from the first subband bottom by taking into account the shift due to impurities in a prescribed electric field; $n_{imp} = 5 \times 10^{10}$ cm^{-2}; $U_d = 58.93$ meV.

lists the basic characteristics of resonances (amplitude j^{res}, full width at half-height Γ, and the ratio of the left-hand half-width at half-height Γ_- to the right-hand half-width at half-height Γ_+) corresponding to the plotted curves.

The following trends exhibited by these resonant-tunneling profiles are of interest.

First, the computed resonance widths are much greater than those based on the uncertainty relation and relaxation times given by the Fermi rule [11, 13, 14] not only for the first-order resonances, but also for higher order resonances. In particular, the tunneling-resonance widths predicted for the structure with $n_{imp} = 1.75 \times 10^{10}$ cm^{-2} by using the results of [11–14] are not greater than 2 meV. Second, the profiles computed for resonances of all orders are characterized by a substantial asymmetry.

According to a direct numerical analysis, optical-phonon- and impurity-assisted scattering processes are dominant in structures of the type considered here, whereas the contributions due to scattering by acoustic phonons and interface roughness are comparatively weak. For this reason, the first two scattering mechanisms are discussed here in some detail. Regarding optical-phonon-assisted scattering, the following remarks should be made. First, the contribution of intersubband transitions of this type must be taken into account even if the intersubband gap is smaller than the optical-phonon energy. Second, note that optical-phonon-assisted intrasubband scattering weakly contributes to the tunneling resonance peak, as compared to intersubband scattering, but plays an important role in resonance tail effects.

A more complicated and interesting pattern is characteristic of scattering by impurities. Figure 3 and 4 show, respectively, the contributions of the impurity-assisted scattering in the first subband to the imaginary part of self-energy and the corresponding partial densities of states calculated as functions of planar energy for several values of total energy. These graphs demonstrate that impurity-assisted intrasubband scattering

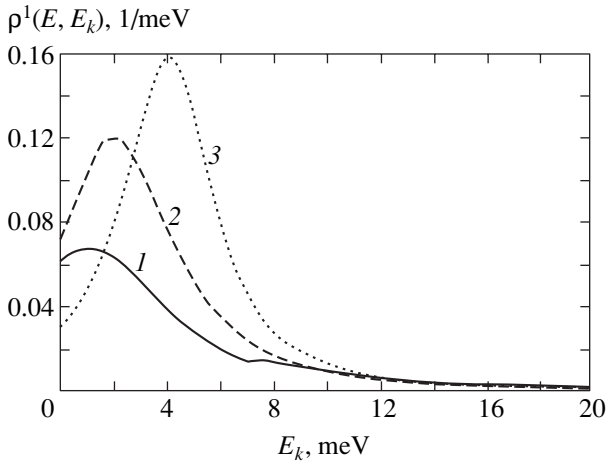

Fig. 4. Partial density of states in a well for the first subband versus the planar energy $E_{\mathbf{k}} = \hbar^2\mathbf{k}^2/2m_w$ for the total energies $E = 0$ (*1*), 2 (*2*), and 4 meV (*3*). The total energies are measured from the first-subband bottom by taking into account the shift due to impurities in a prescribed electric field; $n_{imp} = 5 \times 10^{10}$ cm^{-2}; $U_d = 58.93$ meV.

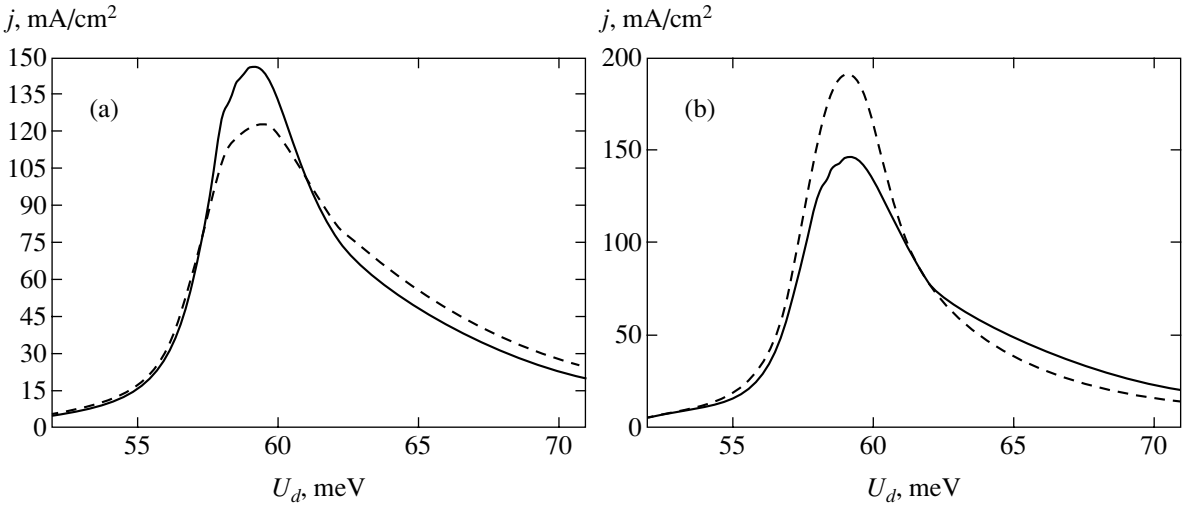

Fig. 5. Tunneling current density between adjacent quantum wells versus voltage drop per period for a $GaAs/Al_{0.3}Ga_{0.7}As$ superlattice in the vicinity of the third-order resonance calculated with (solid curve) and without (dashed curve) allowance for the electric-field effect on intrasubband scattering (a) and thresholds of elastic processes (b); $n_{imp} = 1.75 \times 10^{10}$ cm^{-2}.

leads to considerable uncertainty in one-electron energy. The corresponding contributions to the imaginary part of the self-energy strongly depend on the total energy E. This dependence manifests itself both in the calculated resonance amplitude, width, and shape, being primarily responsible for the asymmetry predicted for high-order resonances. It must be taken into account in calculating the tunneling current density. When the Fermi rule is used, this dependence is ignored and the imaginary part of the contribution of intrasubband scattering to self-energy is calculated much less accurately as a function of planar energy. (It is obvious that the characteristic broadening of one-electron states based on this rule is even less accurate.)

Another effect of particular importance for transport in structures with small size-quantization energies and tunneling into high-lying subbands is the dependence of the states quasi-localized in wells on electric field.

In addition to the electric-field dependence of both energy states in wells and tunneling matrix elements (demonstrated in [8–10]), the resonant-tunneling profiles for structures with small size-quantization energies are affected by the electric-field dependence of scattering processes. Figure 5a shows the resonant-tunneling profiles calculated with and without allowance for the electric-field effect on scattering (the electric-field dependence of energy states in wells and tunneling matrix elements is taken into account in both cases). The corresponding resonance characteristics are listed in Table 2. It is obvious that the electric-field dependence of scattering processes modifies the tunneling-resonance width and asymmetry, as well as its amplitude (albeit to a lesser degree).

Table 2

	Without allowance for the electric-field effect on scattering	With allowance for the electric-field effect on scattering
U_d^{res}, meV	59.34	59.14
j^{res}, mA/cm^2	123.06	146.17
Γ, meV	7.26	5.20
Γ_-/Γ_+	0.47	0.60

We found that the electric-field dependence of both optical-phonon-assisted scattering and impurity-assisted intrasubband scattering is important and must be taken into account. (Note that the former was shown to be substantial for structures with small size-quantization energies in [11].) The mathematical apparatus presented in Section 3 provides an adequate tool for modeling the dependence of impurity-assisted intrasubband scattering on electric field. This is illustrated by Fig. 5a, which shows the tunneling current density calculated with and without allowance for the electric-field effect on scattering.

Furthermore, our analysis of scattering by impurities leads to the following important observation. Calculations of the screened potential (as in the case of wide quantum wells [13, 14]) are commonly based on the Thomas–Fermi model, which does not take into account the threshold behavior of elastic processes. However, we found that this behavior strongly affects the resonant-tunneling profiles obtained for structures with small size-quantization energies, as demonstrated in Fig. 5b by the third-order tunneling-resonance spectra calculated with and without allowance for the thresholds of elastic processes. The corresponding res-

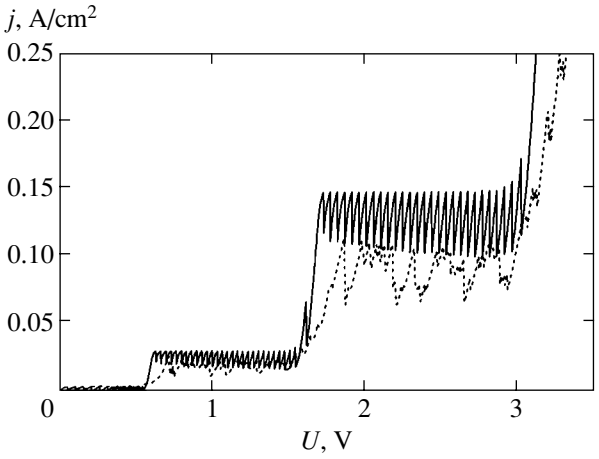

Fig. 6. Comparison of the current–voltage characteristic calculated by using the resonant-tunneling profile for $n_{imp} = 1.75 \times 10^{10}$ cm^{-2} shown in Fig. 2 (solid curve) with a measured one (dashed curve).

onance characteristics are listed in Table 3. It is obvious that the threshold behavior of elastic processes manifests itself in the tunneling-resonance amplitude, width, and asymmetry.

The importance of the thresholds of elastic processes is clear from Fig. 3. The trends illustrated here are explained by the behavior of polarization operator (22) in the vicinity of the thresholds of elastic processes, above which a drastic decrease in screening is observed. The resulting high scattering intensity and the rapid decrease in the threshold value of planar energy with increasing total energy substantially modify the resonant-tunneling profile.

We also revealed an unusual dependence of the characteristics of a tunneling resonance on its order. The difference between resonances of order higher than one is mainly attributed to the difference in tunneling matrix elements between these resonances and in intensity of optical-phonon-assisted scattering between subbands. However, our numerical results suggest that the difference between resonances of different orders is at least equally due to impurity-assisted intrasubband scattering. (Note that scattering by impurities, interface roughness, and acoustic phonons are very weak.) This is clear from the substantial variation in both width and asymmetry of resonances of all orders with impurity concentration (see Fig. 2) and the fact that the extent of the variation depends on the order of a resonance.

Table 3

	Without allowance for the thresholds of elastic processes	With allowance for the thresholds of elastic processes
j^{res}, mA/cm^2	191.72	146.17
Γ, meV	4.05	5.20
Γ_-/Γ_+	0.84	0.60

We used the tunneling-resonance spectra obtained in this study and a discrete model of resonant tunneling transport through superlattices [3] to calculate the current–voltage characteristic of the GaAs/AlGaAs superlattice compared with a measured one in Fig. 6. It is clear that the numerical and experimental results are in good agreement.

Even though the present analysis is focused on tunneling transport through superlattices with small size-quantization energies driven by electric field in the absence of other external factors, the present model can be applied to a broader class of problems arising in studies of transport in weakly coupled superlattices. The general part of the proposed theory, as well as certain ideas and methods used in solving problems of this kind, can be employed to examine the dependence of transport on various time-independent factors, such as nonuniform electric field, magnetic field, etc.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 03-02-16663.

REFERENCES

1. J. Faist, F. Capasso, D. Sirtori, *et al.*, Science **264**, 553 (1994).
2. J. Kastrup, H. T. Grahn, K. Ploog, *et al.*, Appl. Phys. Lett. **65**, 1808 (1994).
3. Yu. A. Mityagin and V. N. Murzin, Pis'ma Zh. Éksp. Teor. Fiz. **64**, 146 (1996) [JETP Lett. **64**, 155 (1996)].
4. A. Wacker, F. Prengel, and E. Scholl, in *Proceedings of 22nd International Conference on the Physics of Semiconductors*, Ed. by D. J. Lockwood (World Sci., Singapore, 1995), p. 1075.
5. H. T. Grahn, J. Kastrup, K. Ploog, *et al.*, Jpn. J. Appl. Phys. **34**, 4526 (1995).
6. J. Kastrup, R. Klann, T. H. Grahn, *et al.*, Phys. Rev. B **52**, 13761 (1995).
7. A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskiĭ, *Methods of Quantum Field Theory in Statistical Physics*, 2nd ed. (Dobrosvet, Moscow, 1988; Prentice-Hall, Englewood Cliffs, N.J., 1963).
8. A. Wacker, in *Theory of Transport Properties of Semiconductor Nanostructures*, Ed. by E. Schöll (Chapman and Hall, London, 1998), p. 321.
9. A. Wacker, A.-P. Jauho, S. Zeuner, and S. J. Allen, Phys. Rev. B **56**, 13268 (1997).
10. A. Wacker and A.-P. Jauho, Phys. Scr. **69**, 321 (1997).

11. R. Ferreira and G. Bastard, Phys. Rev. B **40**, 1074 (1989).
12. V. N. Murzin and Yu. A. Mityagin, Usp. Fiz. Nauk **169**, 464 (1999) [Phys. Usp. **42**, 396 (1999)].
13. V. A. Chuenkov, V. N. Murzin, Yu. A. Mityagin, and L. Yu. Shchurova, Izv. Ross. Akad. Nauk, Ser. Fiz. **65**, 264 (2001).
14. V. N. Murzin, Yu. A. Mityagin, V. A. Chuenkov, *et al.*, Physica E (Amsterdam) **7**, 58 (2000).
15. W. A. Harrison, *Solid State Theory* (McGraw-Hill, New York, 1970; Mir, Moscow, 1972).
16. G. Bastard, *Wave Mechanics Applied to Semiconductor Heterostructures* (Physique, Les Ulis, 1988), p. 63.
17. W. Kohn, Phys. Rev. **115**, 809 (1959).
18. R. F. Kazarinov and R. A. Suris, Fiz. Tekh. Poluprovodn. (Leningrad) **6**, 148 (1972) [Sov. Phys. Semicond. **6**, 120 (1972)].
19. G. D. Mahan, *Many-Particle Physics* (Plenum, New York, 1990).
20. S. M. Goodnick, D. K. Ferry, C. W. Wilmsen, *et al.*, Phys. Rev. B **32**, 8171 (1985).
21. I. Dharssi and P. N. Butcher, J. Phys. C: Condens. Matter **2**, 4629 (1990).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 3, 2004, pp. 633–639.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 3, 2004, pp. 727–733.

SOLIDS
Electronic Properties

Nonlinear Cyclotron–Impurity Resonance in Semiconductors

V. A. Margulis
Ogarev State University, Saransk, 430000 Russia
e-mail: theorphysics@mrsu.ru
Received March 16, 2004

Abstract—The nonlinear absorption of electromagnetic radiation by electrons in a quantizing magnetic field is investigated. The inclusion of multiphoton processes is shown to result in additional peaks in the absorption curve. The shape and arrangement of these peaks were found. The absorption is shown to depend on the electric field strength in the electromagnetic wave nonlinearly and nonmonotonically. The results obtained are in good agreement with experimental data. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Several different types of high-frequency resonance transitions attributable to photon absorption are possible for electrons in the conduction band of a semiconductor placed in a quantizing magnetic field **B**. A cyclotron resonance arises at the radiation frequency $\omega = \omega_c$, and cyclotron–phonon resonances (CFR) accompanied by the emission or adsorption of optical phonons with frequency ω_0 arise at the frequencies $\omega = n\omega_c \pm \omega_0$ (n = 0, 1, 2, …). In particular, cyclotron–phonon transitions can occur with a spin flip or can lead to a transition between the valleys in multivalley semiconductors. A comprehensive overview of the theoretical and experimental works in this area performed before 1978 is given in [1].

The scattering of electrons by impurities can also lead to resonance transitions at the harmonics of the cyclotron frequency ω_c or, more specifically, at $\omega = n\omega_c$ [2, 3]. In these papers, the resonance absorption was investigated by using the randomly arranged pointlike potentials that modeled neutral impurities. The absorption coefficient was calculated in [2, 3] by a method [4] based on perturbation theory. However, using perturbation theory for δ-like potentials in a dimension larger than unity is unacceptable [6].

When the electromagnetic field is strong, apart from single-photon processes, multiphoton processes (nonlinear CFR) are possible. This effect was considered in [7], where the resonance frequencies were shown to be $s\omega = n\omega_c \pm \omega_0$ (s, n = 0, 1, 2, …).

Note that the resonance transitions mentioned above are also possible in the case of scattering by ionized impurities. The transitions of electrons with the absorption of photons and the scattering by such impurities can occur both with and without electron capture by an impurity. Below, we consider only the transitions of the first type. Such cyclotron–impurity transitions in various semiconductors have been extensively studied in the linear (in field) approximation both theoretically and experimentally [8–17]. The two-photon cyclotron–impurity resonance was also studied experimentally [9].

The goal of this work is to investigate the nonlinear resonance absorption of electromagnetic radiation by electrons of the conduction band attributable to scattering by ionized impurities.

We assume that all impurities are identical and are randomly arranged in the sample. If the mean separation between the impurities is much larger than the thermal electron wavelength $\lambda_T = \hbar/\sqrt{2m^*T}$ (a nondegenerate semiconductor) or the Fermi electron wavelength $\lambda_F = \hbar/\sqrt{2m^*\varepsilon_F}$ (a degenerate semiconductor), then the probability of the scattering by N_i scatterers averaged over the impurity positions is equal to the probability of the scattering by one scatterer multiplied by the number of scatterers N_i.

For an impurity located at the coordinate origin, the simplest form of the screened potential energy of the electron–impurity interaction is well known:

$$U(r) = \frac{Ze^2}{\varepsilon r}e^{-\kappa r}. \qquad (1)$$

Here, $\kappa = 1/r_0$, r_0 is the screening radius, Ze is the impurity charge, and ε is the dielectric constant. For the subsequent analysis, it is convenient to represent this potential energy as the Fourier expansion [18]

$$U(\mathbf{r}) = \frac{4\pi Ze^2}{\varepsilon V_0}\sum_{\mathbf{q}}\frac{1}{q^2+\kappa^2}\exp(i\mathbf{q}\cdot\mathbf{r}), \qquad (2)$$

where V_0 is the normalization volume. Let us introduce the quantity $C_{\mathbf{q}} = 4\pi Ze^2/\varepsilon V_0(q^2 + \kappa^2)$. The potential

energy (2) then takes the form

$$U(\mathbf{r}) = \sum_{\mathbf{q}} C_{\mathbf{q}} \exp(i\mathbf{q}\cdot\mathbf{r}).$$

According to the simple screening theory [19], the screening radius for a nondegenerate semiconductor does not depend on magnetic field and is equal to the classical Debye radius. Below, when considering the dependence of absorption on the external magnetic field, κ is assumed to be independent of B. The nonlinear cyclotron–impurity absorption coefficient $\Gamma(\omega)$ can be calculated in the first order of perturbation theory in electron–impurity perturbation by a method similar to that used in [7]. Next, we assume that the magnetic length $l_B = \sqrt{c\hbar/eB}$ is much larger than the lattice constant; hence, we can use the approximation of an electron effective mass [20] that, for simplicity, is taken to be isotropic. Let us assume that the magnetic field $\mathbf{B} \parallel z$ is a quantizing one, the photon energy is $\hbar\omega \gg T$, and the collisional width of the electron levels $\hbar/\tau$ is small compared to the temperature T and $\hbar\omega$. Here, τ is the relaxation time for the electron momentum at the scatterers.

2. THE HAMILTONIAN AND THE TRANSITION PROBABILITY

Let us consider the interaction of electrons with photons and ionized impurities that induces the transitions between Landau magnetic subbands. The impurity-unperturbed Hamiltonian of an electron in the field of an electromagnetic wave and in a stationary and uniform magnetic field is

$$H_0 = \frac{1}{2m^*}\left(\mathbf{p} - \frac{e}{c}\mathbf{A}\right)^2. \tag{3}$$

The vector potential of the fields is $\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2$, where $\mathbf{A}_1 = \mathbf{a}\cos\omega t$, $\mathbf{a} = \mathbf{E}_0/\omega$, and $\mathbf{A}_2 = (-By, 0, 0)$. Here, E_0 and ω are the strength amplitude and frequency of the variable electric field of the wave, and B is the induction of the stationary magnetic field. The exact solution of the Schrödinger equation with Hamiltonian (3) was found in [5]. The electron–impurity interaction operator V is

$$V = \sum_i U(\mathbf{r} - \mathbf{R}_i). \tag{4}$$

Here, $\mathbf{R}_i$ are the positions of the impurities.

The electron transition probability attributable to the simultaneous action of the impurities and the electromagnetic field that includes multiphoton processes can be determined by using results from [7]. After simple, but rather cumbersome transformations, averaging over the impurity positions, we obtain

$$W_{\alpha\alpha'} = \frac{2\pi N_i}{\hbar^2} \times \sum_{q,s} |c_{\mathbf{q}}|^2 D_{l'l}(q_\perp) |P_s(\mathbf{f})|^2 \delta(p'_x, p_x + \hbar q_x) \tag{5}$$
$$\times \delta(p'_z, p_z + \hbar q_z)\delta\left[\omega_c(l'-l) - s\omega + \frac{\hbar q_z^2}{2m^*} - \frac{p_z q_z}{m^*}\right].$$

Here, $q_\perp = \sqrt{q_x^2 + q_y^2}$,

$$D_{l'l}(q_\perp) = \frac{\exp(-\chi)}{2^l l! 2^{l'} l'!} \times \begin{cases} (2^l l'!)^2 \left(\frac{\kappa}{2}\right)^{l-l'} [L_{l'}^{l-l'}(\chi)]^2, & l \geq l', \\ (2^{l'} l!)^2 \left(\frac{\kappa}{2}\right)^{l'-l} [L_l^{l'-l}(\chi)]^2, & l < l', \end{cases} \tag{6}$$

where we introduced $\chi = \hbar q_\perp^2/2m^*\omega_c$. The function $P_s(\mathbf{q})$ in (5) is [5]

$$P_s(\mathbf{q}) = \sum_{s'=-\infty}^{\infty} \exp\left(\frac{i\pi s'}{2}\right) J_{s'-s}(\alpha_1) J_{s'}(\alpha_2), \tag{7}$$

where

$$\alpha_1 = \frac{e a_z q_z}{m^*\omega} - \frac{e(a_x q_x + a_y q_y)}{m^*}\left(\frac{\omega}{\omega_c^2 - \omega^2}\right),$$
$$a_2 = \frac{e(a_y q_x - a_x q_y)}{m^*}\left(\frac{\omega_c}{\omega_c^2 - \omega^2}\right). \tag{8}$$

Here, $J_s(\alpha)$ is the Bessel function.

If the probability of the s-photon process $W^s_{\alpha\alpha'}$ is given by the formula

$$W_{\alpha\alpha'} = \sum_s W^s_{\alpha\alpha'}, \tag{9}$$

then the absorption coefficient Γ^s for the s-photon process can be written as [7]

$$\Gamma^s = \frac{16\pi\hbar n_0 s}{c\sqrt{\varepsilon(\omega)}a^2\omega} \times \left[1 - \exp\left(-\frac{\hbar\omega}{T}\right)\right] \sum_{\alpha,\alpha'} f(\varepsilon_\alpha) W^s_{\alpha\alpha'}. \tag{10}$$

For a degenerate gas, $f(\varepsilon_\alpha)$ in (10) should be substituted with $f_0(\varepsilon_\alpha)[1 - f_0(\varepsilon_{\alpha'})]$, where $f_0(\varepsilon_\alpha)$ is the Fermi function, n_0 is the electron density, $f(\varepsilon_\alpha)$ is the Boltzmann

distribution normalized to unity, and the factor in parentheses includes the induced photon radiation. Below, we restrict our analysis to the transverse case where the electric component of the electromagnetic radiation field $\mathbf{E} \perp \mathbf{B}$, where $\mathbf{B}$ is the induction vector of the stationary and uniform magnetic field. Note that the case of $\mathbf{E} \parallel \mathbf{B}$ is of no interest at the resonance points under consideration (when the frequency detuning $\Delta\omega = \omega_c(l - l') + s\omega = 0$), because the argument of the Bessel function in (5) is proportional to q_z in this case. As follows from the asymptotics of J_s, $\Gamma^s(\Delta\omega = 0) = 0$ at low values of the argument, i.e., at $\Delta\omega = 0$.

If we assume that $E = E_x$ in the transverse case considered below, then

$$\sqrt{\alpha_1^2 + \alpha_2^2} = \frac{eaq_\perp}{m^*(\omega_c^2 - \omega^2)}\sqrt{\omega_c^2\cos^2\varphi + \omega^2\sin^2\varphi}, \quad (11)$$

where $\tan\varphi = q_y/q_x$.

Using the Graf summation formula for the Bessel function in (5), we obtain

$$W^s_{\alpha\alpha'} = \frac{2\pi N_i}{\hbar^2}\sum_{\mathbf{q}}|c_\mathbf{q}|^2 D_{l'l}(q_\perp)$$
$$\times\left|J_s\left(\sqrt{\alpha_1^2 + \alpha_2^2}\right)\right|^2\delta(p'_x, p_x + \hbar q_x) \quad (12)$$
$$\times\delta(p'_z, p_z + \hbar q_z)\delta\left(\Delta\omega - \frac{\hbar q_z^2}{2m^*} + \frac{p_z q_z}{m^*}\right).$$

Substituting (12) into (10) yields

$$\Gamma^s = \frac{4\pi^2 V_0 n_0 s N_i}{c\hbar\sqrt{\varepsilon(\omega)}a^2\omega}\left(1 - \exp\left(-\frac{\hbar\omega}{T}\right)\right)$$
$$\times\sum_{\alpha, l'}\int dq_\parallel\int dq_\perp q_\perp|C_\mathbf{q}|^2 D_{l'l}(q_\perp) \quad (13)$$
$$\times f(\varepsilon_\alpha)\delta\left(\Delta\omega - \frac{\hbar^2 q_\parallel^2}{2m^*} + \frac{p_\parallel q_\parallel}{m^*}\right)$$
$$\times\int_0^{2\pi}\left|J_s\left(\frac{eaq_\perp}{m^*}\frac{\sqrt{\omega_c^2\cos^2\varphi + \omega^2\sin^2\varphi}}{\omega_c^2 - \omega^2}\right)\right|^2 d\varphi,$$

where $q_z \equiv q_\parallel$.

The single-photon absorption coefficient can be derived from (13) by passing to the limit, setting $s = 1$, and using the inequality for a weak field E_x:

$$\frac{eaq_\perp}{m^*}\frac{\sqrt{\omega_c^2\cos^2\varphi + \omega^2\sin^2\varphi}}{\omega^2 - \omega_c^2} \ll 1. \quad (14)$$

Taking the asymptotics of the Bessel functions at low values of the argument, we obtain from (13)

$$\Gamma^1 = \frac{(2\pi)^3 e^2 V_0 n_0 N_i}{4c\hbar\sqrt{\varepsilon(\omega)}m^{*2}\omega}\left[\frac{1}{(\omega - \omega_c)^2} + \frac{1}{(\omega + \omega_c)^2}\right]$$
$$\times\sum_{\alpha, l'}\int dq_\parallel\int dq_\perp q_\perp^3 \quad (15)$$
$$\times|C_\mathbf{q}|^2 D_{ll'}(q_\perp) f(\varepsilon_\alpha)\delta\left(\Delta\omega - \frac{\hbar q_\parallel^2}{2m^*} + \frac{p_\parallel q_\parallel}{m^*}\right).$$

The same result is obtained when considering the transitions in the second order in electron–impurity and electron–photon perturbations by the method described in [4].

When the inequality opposite to (14) holds, a highly nonlinear case considered below takes place. Using the asymptotics of the Bessel functions at high values of the argument, the integral in (13) over the angle φ can be estimated as

$$\int_0^{2\pi}|J_s|^2 d\varphi \approx \frac{2m^*}{\pi eaq_\perp|\omega_c^2 - \omega^2|}\int_0^{2\pi}\frac{\cos^2[eaq_\perp\sqrt{\omega^2\sin^2\varphi + \omega_c^2\cos^2\varphi}/(m^*|\omega_c^2 - \omega^2|) - \pi s/2 - \pi/4]}{\sqrt{\omega^2\sin^2\varphi + \omega_c^2\cos^2\varphi}}d\varphi. \quad (16)$$

Averaging the square of the cosine of the large argument in (16), i.e., substituting 1/2 for it, yields

$$\int_0^{2\pi}|J_s|^2 d\varphi \approx \begin{cases}\dfrac{8m^*|\omega_c^2 - \omega^2|}{\pi eaq_\perp\omega}K\left(\dfrac{\sqrt{|\omega_c^2 - \omega^2|}}{\omega}\right), & \text{if } \omega > \omega_c, \\ \dfrac{8m^*|\omega_c^2 - \omega^2|}{\pi eaq_\perp\omega_c}K\left(\dfrac{\sqrt{|\omega_c^2 - \omega^2|}}{\omega_c}\right), & \text{if } \omega_c > \omega.\end{cases} \quad (17)$$

Here $K(x)$ is the complete elliptic integral of the first kind. Substituting (17) into (13) yields

$$\Gamma^s = A_s(\omega)\sum_{\alpha, l'}\int dq_\parallel\int dq_\perp|C_\mathbf{q}|^2 D_{ll'}(q_\perp) f(\varepsilon_\alpha)$$
$$\times\delta\left(\Delta\omega - \frac{\hbar q_\parallel^2}{2m^*} + \frac{p_\parallel q_\parallel}{m^*}\right), \quad (18)$$

where

$$A_a(\omega) = \begin{cases} \dfrac{16\pi V_0 n_0 N_i m^* |\omega_c^2 - \omega^2| s}{ec\hbar\sqrt{\varepsilon(\omega)}a^3\omega^2} K\left(\dfrac{\sqrt{|\omega_c^2 - \omega^2|}}{\omega}\right)\left[1 - \exp\left(-\dfrac{\hbar\omega}{T}\right)\right], & \text{if } \omega > \omega_c, \\ \dfrac{16\pi V_0 n_0 N_i m^* |\omega_c^2 - \omega^2| s}{ec\hbar\sqrt{\varepsilon(\omega)}a^3\omega\omega_c} K\left(\dfrac{\sqrt{|\omega_c^2 - \omega^2|}}{\omega_c}\right)\left[1 - \exp\left(-\dfrac{\hbar\omega}{T}\right)\right], & \text{if } \omega < \omega_c. \end{cases} \quad (19)$$

3. THE SHAPE AND INTENSITY OF THE RESONANCE PEAKS

Next, let us consider the integral over $q_\perp$:

$$B(q_\parallel) = \int_0^\infty dq_\perp |C_{\mathbf{q}}|^2 D_{ll'}(q_\perp). \quad (20)$$

Because of the δ function in (18), the absorption has a singularity at $\Delta\omega = 0$. The behavior of this singularity can be analyzed by taking $\Delta\omega \ll \omega_c$. The domain of integration whose sizes are determined by the largest of the quantities k_T or $\sqrt{(2m^*\Delta\omega)}/\hbar$ will then be significant in the integral over $p_\parallel$ and $q_\parallel$ in (18) due to the factors $f(\varepsilon_\alpha)$ and δ. However, the values of both these quantities are much smaller than the magnetic length l_B. Therefore, to study the behavior of the singularity, we can assume $B(q_\parallel)$ to be a constant function and set $q_\parallel = 0$ in it. If the impurity concentration and the temperature are such that $\kappa = (\varepsilon T/4\pi n_0 e^2)^{1/2}$ satisfies the inequality $\kappa \ll 1/l_H$ (the condition for a fairly strong magnetic field), then we may disregard the screening in (20) and set $\kappa = 0$ in it. In this case, $B(q_\parallel)$ takes the form

$$B(q_\parallel) \approx \left(\frac{4\pi Ze^2}{V_0}\right)^2 \int_0^\infty dq_\perp \frac{1}{q_\perp^4} D_{ll'}(q_\perp). \quad (21)$$

The integral in (21) can be calculated by using (21) and expression (6) for $D_{ll'}(q_\perp)$. Let us first reduce it to the form

$$B(q_\parallel) = \left(\frac{4\pi Ze^2}{\varepsilon V_0}\right)^2 \frac{l_B^3}{4\sqrt{2}} \frac{l!}{l'!} \times \int_0^\infty e^{-x} x^{l'-l-5/2} [L_l^{l'-l}(x)]^2 dx, \quad l' \geq l. \quad (22)$$

For $l' < l$, we must substitute $l' \longleftrightarrow l$ in (22). Using (22) to determine $B(q_\parallel)$, we then obtain for $l - l' \geq 2$

$$B(q_\parallel) = \left(\frac{4\pi Ze^2}{\varepsilon V_0}\right)^2 \frac{l_B^3 l!}{l'!} \frac{\Gamma(l'-l-3/2)\Gamma(l'+1)}{\Gamma(l'-l+1)(l!)^2} \times \left\{ \frac{d^l}{dx^l}\left[\frac{F\left(\dfrac{l'-l-3/2}{2}, \dfrac{l'-l-1/2}{2}; l'-l+1; \dfrac{4x}{(1+x)^2}\right)}{(1+x)^{l'-l-3/2}(1-x)^{5/2}} \right] \right\}_{x=0}, \quad (23)$$

where $\Gamma(x)$ is the Euler function and $F(\alpha, \beta, \gamma, \delta)$ is the hypergeometric function.

In the most interesting case where $l = 0$ and $l' > 1$, expressions (23) greatly simplifies:

$$B(q_\parallel) = \left(\frac{4\pi Ze^2}{\varepsilon V_0}\right)^2 \frac{l_B^3}{2\sqrt{2}} \Gamma\left(l' + \frac{1}{2}\right). \quad (24)$$

At $l = 0$ and $l' = 1$, integral (22) diverges at the lower limit. However, if we take $q_\parallel$, $\kappa \neq 0$, then we obtain from (22) [23]

$$B(q_\parallel) = \left(\frac{4\pi Ze^2}{\varepsilon V_0}\right)^2 \frac{l_B^2 \Gamma(3/2)}{2\sqrt{q_\parallel^2 + \kappa^2}} \psi\left[\frac{3}{2}, \frac{1}{2}; \frac{l_B^2(q_\parallel^2 + \kappa^2)}{2}\right], \quad (25)$$

where $\psi(\alpha, \beta; x)$ is the degenerate hypergeometric function. Since the denominator is small, this term mainly contributes to the absorption intensity. As follows from (22)–(25), $B(q_\parallel)$ can always be represented as

$$B(q_\parallel) = \left(\frac{4\pi Ze^2}{\varepsilon V_0}\right)^2 M_{l'l}. \quad (26)$$

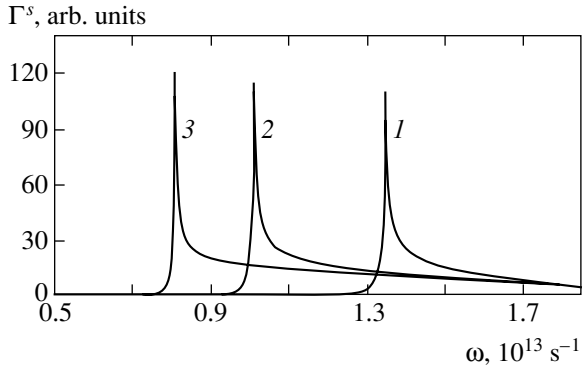

Fig. 1. Partial absorption coefficients versus radiation frequency for $\omega_c = 2.02 \times 10^{13}$ s^{-1}, $T = 10$ K, $l = 2$, and $l' = 0$; curves *1*, *2*, and *3* correspond to $s = 3$, 4, and 5, respectively.

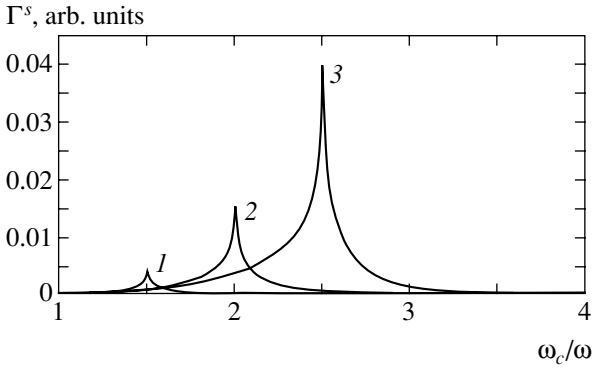

Fig. 2. Partial absorption coefficients versus magnetic field for $\omega = 5 \times 10^{12}$ s^{-1}, $T = 30$ K, $l = 2$, and $l' = 0$; curves *1*, *2*, and *3* correspond to $s = 3$, 4, and 5, respectively.

It then follows from (18) that the partial absorption coefficients $\Gamma_{l'l}$ ($\Gamma^s = \sum_{l,l'} \Gamma^s_{l'l}$) near the resonance points are

$$\Gamma^s_{l'l} = A_s(\omega)\left(\frac{4\pi Ze^2}{\varepsilon V_0}\right)^2 \sum_{p_x, p_z} M_{l'l} f(\varepsilon_\alpha) \times \int dp'_\parallel \delta\left(\hbar\Delta\omega + \frac{p_\parallel^2}{2m^*} - \frac{p_\parallel'^2}{2m^*}\right), \quad (27)$$

where $p_z = p_\parallel$. We find from (27) that

$$\Gamma^s_{l'l} = 2m^* A_s(\omega)\left(\frac{4\pi Ze^2}{\varepsilon V_0}\right)^2 M_{l'l} \int_0^\infty \frac{f(\varepsilon_\parallel + \varepsilon_l)d\varepsilon}{\sqrt{\varepsilon_\parallel(\varepsilon_\parallel + \hbar\Delta\omega)}}, \quad (28)$$

where $\varepsilon_\parallel = p_\parallel^2/2m^*$ and $\varepsilon_l = \hbar\omega_c(l + 1/2)$. According to [4], $f(\varepsilon_\parallel + \varepsilon_l)$ is given by

$$f(\varepsilon_\parallel + \varepsilon_l) = \frac{2\sinh(\hbar\omega_c/T)}{\sqrt{2\pi m^* T}} \exp\left[-\frac{\varepsilon_\parallel + \varepsilon_l}{T}\right]. \quad (29)$$

We find from (28) and (29) that

$$\Gamma^s_{l'l} = \frac{4\sinh(\hbar\omega_c/T)m^*}{\sqrt{2\pi m^* T}}\left(\frac{4\pi Ze^2}{\varepsilon V_0}\right)^2 A_s(\omega) \times \exp\left\{\left[\hbar\Delta\omega - \hbar\omega_c\left(l + \frac{1}{2}\right)\right]/T\right\} K_0\left(\frac{\hbar|\Delta\omega|}{T}\right). \quad (30)$$

Here, $K_0(x)$ is the Macdonald function. Since the Macdonald function has a logarithmic singularity at $x = 0$, there is a resonance in the absorption of electromagnetic radiation when the condition $\Delta\omega = 0$ is satisfied. We determine the resonance frequencies from this condition:

$$\omega_r^s = \frac{l' - l}{s}\omega_c. \quad (31)$$

As follows from (30), the behavior of the singularity in the case considered is the same as that in CFR. Consequently, the conclusion reached in [4] about the behavior of the curve to the left and to the right of the points where $\Delta\omega = 0$ is also valid in our case. To the right of the resonance points where $\hbar\Delta\omega \gg T$, the absorption depends on $\Delta\omega$ as $1/\sqrt{\Delta\omega}$, transforming into $\ln\hbar|\Delta\omega|/T$ at $\hbar\Delta\omega \ll T$, while to the left, the singularity is also logarithmic at $\Delta\omega < 0$ and $\hbar|\Delta\omega| \ll T$, but the exponential decrease $|\Delta\omega|^{-1/2}\exp[-\hbar|\Delta\omega|/T]$ is superimposed on the square-root singularity far at $\Delta\omega < 0$ and $\hbar|\Delta\omega| \gg T$. Thus, the resonance peaks are asymmetric; i.e., the right wing of the peak is flatter than the left wing. The same result also shows up in the experimental data [9]. Interestingly, the dependence $\Gamma^s(\omega_c)$ differs from $\Gamma^s(\omega)$ in that the left wing in them is flatter than the right wing (Figs. 1 and 2). This is attributable to the factor $\sinh(\hbar\omega_c/T)$ in (30). As follows from (31), an *s*-photon resonance takes place at fractional multiples of the cyclotron frequency. For single-photon absorption ($s = 1$), a resonance arises at integer multiples of the cyclotron frequency, i.e., at the harmonics of the cyclotron frequency. Comparing expression (15) for single-photon absorption with (30), we see that the shape of the resonance peaks in the absorption curve and the behavior of the singularity at the resonance point in the nonlinear case are the same as those in the linear case. The same conclusion can be reached with regard to the dependence of the amplitude of the resonance peaks in the absorption curve on the temperature and the magnetic field. The dependence of the absorption coefficient on the number of photons involved in the transitions is determined by the function $A_s(\omega)$. As follows from (19), this dependence is mainly defined as s^{-1}. Therefore, the amplitude of the photon repetitions of the main ($s = 1$) peak decreases with increasing *s*. This estimate agrees well with the experimental data [9] for the two-photon cyclotron–impurity resonance. We see from (31) that these repetitions are arranged not equi-

distantly in the curve $\Gamma^s(\omega)$, but in such a way that the separation between the nearest peaks is $\omega_r^s - \omega_r^{s+1} \propto [s(s+1)]^{-1}$. Thus, the separations between the nearest repetitions of the main peak decrease with increasing s.

4. CONCLUSIONS

It is interesting to compare the amplitudes of the nonlinear and linear resonance peaks: in the main case, we obtain the following estimate from (15) and (19) using (24):

$$\frac{(\Gamma_{l'0}^1)^{\text{nonlin}}}{(\Gamma_{l'0})^{\text{lin}}} \propto \left(\sqrt{\frac{m^*\hbar\omega_c^3}{eE_0}}\right)^3. \tag{32}$$

Let us compare the intensities of the peaks of the linear cyclotron–impurity resonance (CIR) and the cyclotron–phonon resonance (CFR). To this end, we write $\Gamma_{l'0}^1$ as

$$\Gamma_{l'0}^1(\text{CIR}) = \left(\frac{4\pi Ze^2}{V_0}\right)\frac{l_H^3\Gamma(l'+1/2)}{2\sqrt{2}}$$

$$\times\frac{(2\pi)^3e^2V_0n_0N_i}{4c\hbar\sqrt{\varepsilon}m^*\omega}\left[\frac{1}{(\omega-\omega_c)^2}+\frac{1}{(\omega+\omega_c)^2}\right] \tag{33}$$

$$\times\frac{\sinh(\hbar\omega_c/T)m^*}{\sqrt{2\pi m^*T}}\exp\left[\hbar\Delta\omega-\hbar\omega_c\left(l'+\frac{1}{2}\right)\right]K_0\left(\frac{\hbar|\Delta\omega|}{T}\right).$$

We take $\Gamma_{l'0}^{\text{CFR}}$ for LO phonons from [4]. In that case,

$$\frac{\Gamma^{\text{CIR}}}{\Gamma^{\text{CFR}}} \approx \frac{(4\pi)^{3/2}(l'-1/2)Z^2e^2l_B^2}{\hbar\omega_0}\left(\frac{N_i}{\varepsilon V_0}\right). \tag{34}$$

For $B = 10$ T and $N_i \approx 10^{15}$ cm^{-3}, $\Gamma^{\text{CIR}}/\Gamma^{\text{CFR}} \sim 4$. Thus, under the experimental conditions [24], the intensity of the linear CIR peaks is much higher than that of the CFR peaks. A similar result was also obtained in the experiment [24]. It was hypothesized in [2] that the peaks observed in [24] at the harmonics of the cyclotron resonance are attributable to the scattering by ionized impurities. The above estimates are consistent with this hypothesis.

For the nonlinear CIR to be observed, a certain condition for the field strength of the electromagnetic wave must be satisfied. It follows from (11) that the nonlinearity condition for the effect at $\omega \sim \omega_c$ is

$$eE_0q_\perp/m^*\omega_c \gtrsim 1. \tag{35}$$

Assuming that $q_\perp \sim l_B^{-1}$, we obtain $eE_0l_B \gtrsim \hbar\omega_c \gg T$ from (35). This inequality is known [25] to be the criterion for a significant dependence of the electron transition probability in a magnetic field on the amplitude of the variable electric field.

Having estimated $\omega_c \sim 2\times10^{13}$ s^{-1} and $q_\perp = 10^5$ cm, we obtain an estimate for $E_0 \sim 10^3$ V cm^{-1} that follows from (35) for the *n*-InSb parameters. The values of $\omega \sim 10^{14}$ s^{-1} and $E_0 \sim 10^3$ V cm^{-1} can be obtained by illuminating the sample by the electromagnetic radiation of a CO_2 laser [26]. This radiation is commonly used to experimentally study the CIR.

To estimate the partial absorption coefficients at maximum, we substitute $\hbar\Delta\omega + i\delta$, where δ is the phenomenological damping constant, for the argument $\hbar\Delta\omega$ in the resonance factor in (30),

$$f(\Delta\omega) = \exp\left\{\frac{\hbar\Delta\omega-\hbar\omega_c(l+1/2)}{T}\right\}K_0\left(\frac{\hbar|\Delta\omega|}{T}\right),$$

that contains a weak (logarithmic) singularity. This factor near the resonance absorption will then take the form

$$\begin{aligned}\text{Re}\,f(\Delta\omega+i\delta) &= (\ln 2 - C)\cos\frac{\delta}{T}\\ &+ \sin\frac{\delta}{T}\arcsin\left(\frac{\delta}{\sqrt{\delta^2+\hbar^2(\Delta\omega)^2}}\right)\\ &- \frac{1}{2}\cos\frac{\delta}{T}\ln\left[\frac{\hbar^2(\Delta\omega)^2+\delta^2}{T^2}\right],\end{aligned} \tag{36}$$

where C is the Euler constant.

The first two terms in this expression make a monotonic contribution to the absorption, while the last term makes a resonance contribution. Let us estimate this factor at the resonance point by taking $\delta = \hbar/\tau_0$, where the relaxation time $\tau_0 \sim 10^{-12}$ s. At $T = 10$ K, we then have $\text{Re}f[\Delta\omega\hbar + i\delta] \sim 1$. For the linear absorption, we obtain $\Gamma_{0l}^{(1)}(\text{CIR}) \sim 0.3$ cm^{-1} at the resonance point. As follows from (32), $\Gamma^{\text{nonlin}}/\Gamma^{\text{lin}} \sim 100$ at $\omega_c \sim 10^{13}$ s^{-1} and $E_0 \sim 10^3$ V cm^{-1} and, hence, $\Gamma^{\text{nonlin}} \gg \Gamma^{\text{lin}}$. However, the absorption decreases with growing field in the region of the nonlinear effect, so that the E_0 dependence of the amplitude of the peaks is nonmonotonic. As follows from (13), this dependence is attributable to the nonmonotonic dependence of the Bessel function on its argument.

For a degenerate gas, to study the resonances in the absorption coefficient at low temperatures, the distribution function may be assumed to be $f_0(\varepsilon_\alpha) = 1$. The absorption coefficient near the resonance points then has the same logarithmic singularity as in the nondegenerate case. This is because the singularity originates from the singularities in the density of the initial and final states and does not depend on the electron gas degeneracy. In the above calculations, we ignored the heating of the electron gas by the electric field of the electromagnetic radiation. The heating may be disre-

garded if $(eE_0 l_H/T)^2\delta \ll 1$, where δ is the scattering inelasticity parameter [25].

REFERENCES

1. R. K. Bakanas, F. G. Bass, and I. B. Levinson, Fiz. Tekh. Poluprovodn. (Leningrad) **12**, 1457 (1978) [Sov. Phys. Semicond. **12**, 863 (1978)].
2. Yu. A. Gurevich, Zh. Éksp. Teor. Fiz. **61**, 1120 (1971) [Sov. Phys. JETP **34**, 598 (1971)].
3. R. I. Rabinovich, Fiz. Tekh. Poluprovodn. (Leningrad) **8**, 91 (1973) [Sov. Phys. Semicond. **8**, 55 (1973)].
4. F. G. Bass and I. B. Levinson, Zh. Éksp. Teor. Fiz. **49**, 914 (1965) [Sov. Phys. JETP **22**, 635 (1965)].
5. V. P. Oleĭnik, Ukr. Fiz. Zh. **13**, 1205 (1968).
6. A. I. Baz', Ya. B. Zel'dovich, and A. M. Perelomov, *Scattering, Reactions and Decays in Nonrelativistic Quantum Mechanics*, 2nd ed. (Nauka, Moscow, 1971; Israel Program for Scientific Translations, Jerusalem, 1966).
7. V. A. Margulis, Fiz. Tekh. Poluprovodn. (Leningrad) **17**, 910 (1983) [Sov. Phys. Semicond. **17**, 571 (1983)].
8. J. Kono, S. Takeyama, H. Yokoi, and N. Miura, Phys. Rev. B **48**, 10909 (1993).
9. W. Bohm, E. Ettlinger, and W. Prettl, Phys. Rev. Lett. **47**, 1198 (1981).
10. M. A. Hopkins, R. J. Nicholas, D. J. Barnes, and M. Brunnell, Phys. Rev. B **39**, 13302 (1989).
11. E. E. H. Shin, P. N. Argyres, and B. Lax, Phys. Rev. Lett. **28**, 1634 (1972).
12. E. E. H. Shin, P. N. Argyres, and B. Lax, Phys. Rev. B **7**, 3572 (1973).
13. J. R. Apel and T. O. Poehler, Phys. Rev. B **4**, 436 (1971).
14. V. K. Arora and M. A. Al-Mass'ari, Phys. Rev. B **23**, 5619 (1981).
15. G. Bastard, J. Mycielski, and C. Rigaux, Phys. Rev. B **18**, 6990 (1978).
16. J. Van Royen, J. De Sitter, and J. T. Devreese, Phys. Rev. B **30**, 7154 (1984).
17. D. Dunn and A. Suzuki, Phys. Rev. B **29**, 942 (1984).
18. J. M. Ziman, *Electrons and Phonons* (Clarendon Press, Oxford, 1960; Inostrannaya Literatura, Moscow, 1962).
19. P. N. Argures and E. N. Adams, Phys. Rev. **104**, 900 (1956).
20. W. Kohn and J. M. Luttiger, Phys. Rev. **108**, 690 (1957).
21. J. M. Luttiger and W. Kohn, Phys. Rev. **97**, 869 (1955).
22. I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products*, 5th ed. (Nauka, Moscow, 1971; Academic, New York, 1980).
23. A. P. Prudnikov, Yu. A. Brychkov, and O. I. Marichev, *Integrals and Series, Elementary Functions* (Nauka, Moscow, 1981; Gordon and Breach, New York, 1986).
24. E. J. Johnson and D. H. Dickey, Phys. Rev. B **1**, 2676 (1970).
25. A. M. Zlobin and P. S. Zyryanov, Usp. Fiz. Nauk **104**, 353 (1971) [Sov. Phys. Usp. **14**, 379 (1971)].
26. Yu. I. Balkareĭ and É. M. Épshteĭn, Zh. Éksp. Teor. Fiz. **63**, 660 (1972) [Sov. Phys. JETP **36**, 349 (1973)].

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 99, No. 3, 2004, pp. 640–646.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 3, 2004, pp. 734–740.

SOLIDS
Electronic Properties

Composite Fermions, Trios, and Quartets in the Fermi–Bose Mixture of Neutral Particles¶

M. Yu. Kagan[a], I. V. Brodsky[a], D. V. Efremov[b], and A. V. Klaptsov[c]
[a]*Kapitza Institute for Physical Problems, Moscow, 119334 Russia*
e-mail: kagan@kapitza.ras.ru
[b]*Technische Universität Dresden Institut für Theoretische Physik, Dresden, 01062 Germany*
[c]*Russian Research Centre "Kurchatov Institute", Moscow, 123182 Russia*
Received March 17, 2004

Abstract—We consider the model of a Fermi–Bose mixture with strong hard-core repulsion between particles of the same sort and attraction between particles of different sorts. In this case, in addition to the standard anomalous averages of the type $\langle b\rangle$, $\langle bb\rangle$, and $\langle cc\rangle$, pairing between fermions and bosons of the type $\langle bc\rangle$ is possible. This pairing corresponds to creation of composite fermions in the system. At low temperatures and equal densities of fermions and bosons, composite fermions are further paired into quartets. At higher temperatures, trios consisting of composite fermions and elementary bosons are also present in the system. Our investigations are important in connection with the recent observation of weakly bound dimers in magnetic and optical dipole traps at ultralow temperatures and with the observation of collapse of a Fermi gas in an attractive Fermi–Bose mixture of neutral particles. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The Fermi–Bose mixture model is currently very popular in connection with different problems in condensed matter physics, such as high-T_c superconductivity, superfluidity in ^{3}He–^{4}He mixtures [1], and fermionic superfluidity in magnetic traps.

In high-T_c superconductivity, this model was first proposed by Ranninger [2, 3] for simultaneous description of high transition temperature and short coherence length of superconductive pairs, on the one hand, and of the presence of a well-defined Fermi surface, on the other.

In this paper, we show that the Fermi–Bose mixture with attractive interaction between fermions and bosons is unstable with respect to the creation of composite fermions $f = bc$. Moreover, for low temperatures and equal densities of fermions and bosons, composite fermions are further paired into $\langle ff\rangle$ quartets. We note that the matrix element $\langle f\rangle = \langle bc\rangle$ is nonzero only for transitions between states $|N_B; N_F\rangle$ and $\langle N_B - 1; N_F - 1|$, where N_B and N_F are the numbers of elementary bosons and fermions, respectively. For the superconductive state, the matrix element $\langle ff\rangle$ is nonzero only for transitions between states $|N_B; N_F\rangle$ and $\langle N_B - 2; N_F - 2|$. Our results are interesting not only for the physics of high-T_c superconductors, but also for Fermi–Bose mixtures in magnetic and optical dipole traps, as well as in optical lattices, where we can easily tune such parameters of the system as the particle density and the sign and strength of the interparticle interaction [4, 5].

2. THEORETICAL MODEL

The model of the Fermi–Bose mixture has the following form on a lattice:

$$H = H_F + H_B + H_{BF},$$

$$H_F = -t_F\sum_{\langle i,j\rangle} c^+_{i\sigma}c_{j\sigma} + U_{FF}\sum_i n^F_{i\uparrow}n^F_{i\downarrow} - \mu_F\sum_{i,\sigma} n^F_{i\sigma},$$

$$H_B = -t_B\sum_{\langle i,j\rangle} b^+_i b_j + \frac{1}{2}U_{BB}\sum_i n^B_i n^B_i - \mu_B\sum_i n^B_i, \quad (1)$$

$$H_{BF} = -U_{BF}\sum_{i,\sigma} n^B_i n^F_{i\sigma}.$$

This is a lattice analog of the standard Hamiltonian considered, for example, in [6]. Here, t_F and t_B are fermionic and bosonic hopping amplitudes, and $c^+_{i\sigma}$, $c_{i\sigma}$, b^+_i and b_i are the fermionic and bosonic creation and annihilation operators; the Hubbard interactions U_{FF} and U_{BB} correspond to hard-core repulsion between particles of the same sort; the interaction U_{BF} corresponds to attraction between fermions and bosons; $W_F = 8t_F$ and $W_B = 8t_B$ are the bandwidths in the two-dimensional case; and finally, μ_F and μ_B are fermionic and bosonic chemical potentials. For the square lattice, the spectra

¶This article was submitted by the authors in English.

of fermions and bosons after the Fourier transformation are given by

$$\xi_{p\sigma} = -2t_F(\cos p_x d + \cos p_y d) - \mu_F$$

for fermions and

$$\eta_p = -2t_B(\cos p_x d + \cos p_y d) - \mu_B$$

for bosons, where d is the lattice constant. In the intermediate coupling case

$$\frac{W_{BF}}{\ln(W_{BF}/T_{0BF})} < U_{BF} < W_{BF},$$

the energy of the bound state is given by

$$|E_b| = \frac{1}{2m_{BF}d^2}\frac{1}{\exp[2\pi/m_{BF}U_{BF}]-1}, \quad (2)$$

where

$$m_{BF} = \frac{m_B m_F}{m_B + m_F}$$

is an effective mass

$$W_{BF} = \frac{4}{m_{BF}d^2},$$

$$T_{0BF} = 2\pi n/m_{BF}.$$

For simplicity, we consider the case of equal densities $n_B = n_F = n$.

We note that in the intermediate coupling case, the binding energy $|E_b|$ between a fermion and a boson is larger than the bosonic and fermionic degeneracy temperatures

$$T_{0B} = \frac{2\pi n_B}{m_B}$$

and

$$T_{0F} = \frac{2\pi n_F}{m_F} \equiv \varepsilon_F,$$

but smaller than the bandwidths W_B and W_F. In this case, pairing of fermions and bosons, $\langle bc\rangle \neq 0$, occurs earlier (at higher temperatures) than both Bose–Einstein condensation of bosons (or bibosons) ($\langle b\rangle \neq 0$ or $\langle bb\rangle \neq 0$) and superconductive pairing of fermions ($\langle cc\rangle \neq 0$). We note that in the case of a very strong attraction $U_{BF} > W_{BF}$, we have the natural result $|E_b| = U_{BF}$, and the effective mass

$$m^*_{BF} = \frac{m_{BF}U_{BF}}{W_{BF}} \gg m_{BF}$$

is additionally enhanced on the lattice [7]. We also note that the Hubbard interactions U_{FF} and U_{BB} satisfy the inequalities

$$U_{FF} = \frac{W_F}{\ln(W_F/|E_b|)},$$

$$U_{BB} > \frac{W_B}{\ln(W_B/|E_b|)}.$$

We now consider the temperature evolution of the system. It is governed by the corresponding Bethe–Salpeter equation. After analytic continuation $i\omega_n \longrightarrow \omega + i0$ [8], the solution of this equation becomes

$$\Gamma(\mathbf{q},\omega) = \frac{-U_{BF}}{1 - U_{BF}\int\frac{d^2p}{(2\pi)^2}\frac{1-n_F(\xi(\mathbf{p}))+n_B(\eta(\mathbf{q}-\mathbf{p}))}{\xi(\mathbf{p})+\eta(\mathbf{q}-\mathbf{p})-\omega-i0}}, \quad (3)$$

where

$$\xi(\mathbf{p}) = \frac{p^2}{2m_F} - \mu_F,$$

$$\eta(\mathbf{p}) = \frac{p^2}{2m_B} - \mu_B$$

are spectra of fermions and bosons at low densities $n_F d^2 \ll 1$ and $n_B d^2 \ll 1$. We note that the temperature factor

$$1 - n_F(\xi(\mathbf{p})) + n_B(\eta(\mathbf{q}-\mathbf{p}))$$

enters the pole of the Bethe–Salpeter equation, in contrast with the factor

$$1 - n_F(\xi(\mathbf{p})) - n_F(\xi(\mathbf{q}-\mathbf{p}))$$

for two-fermion superconductive pairing and

$$1 + n_B(\eta(\mathbf{p})) + n_B(\eta(\mathbf{q}-\mathbf{p}))$$

for two-boson pairing. The pole of the Bethe–Salpeter equation corresponds to the spectrum of the composite fermions,

$$\omega \equiv \xi^*_{\mathbf{p}} = \frac{p^2}{2(m_B + m_F)} - \mu_{\text{comp}}, \quad (4)$$

where

$$\mu_{\text{comp}} = \mu_B + \mu_F + |E_b| \quad (5)$$

is the chemical potential of composite fermions. Composite fermions are well-defined quasi-particles, because the damping of quasi-particles is equal to zero in the case of a bound state ($E_b < 0$), but becomes nonzero and is proportional to E_b in the case of a virtual state ($E_b > 0$). The dynamical equilibrium (boson +

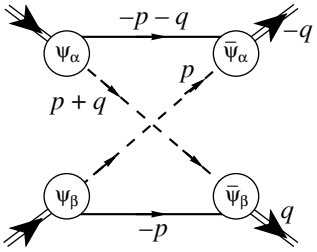

Fig. 1. Skeleton diagram for coefficient b at Ψ^4 in the effective action. Dashed lines correspond to bosons, and solid lines correspond to fermions.

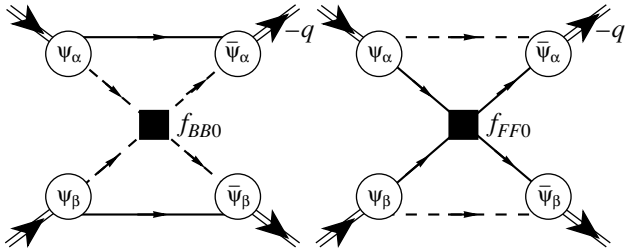

Fig. 2. Corrections to coefficient b containing boson–boson and fermion–fermion interactions.

fermion $\rightleftharpoons$ composite fermion) is governed by the standard Saha formula [9]. In the two-dimensional case, it is

$$\frac{n_B n_F}{n_{\text{comp}}} = \frac{m_{BF}T}{2\pi}\exp\left\{-\frac{|E_b|}{T}\right\}. \tag{6}$$

The crossover temperature T_* is determined, as usual, from the condition that the number of composite fermions is equal to the number of unbound fermions and bosons:

$$n_{\text{comp}} = n_B = n_F = n.$$

This conditions yields

$$T_* \approx \frac{|E_b|}{\ln(|E_b|/2T_{0BF})} \gg \{T_{0B}; T_{0F}\}. \tag{7}$$

We note that in the Boltzmann regime $|E_b| > \{T_{0B}; T_{0F}\}$, we actually deal with the pairing of two Boltzmann particles. Therefore, this pairing does not differ drastically from the pairing of two particles of the same type of statistics. Indeed, if we replace $\mu_B + \mu_F$ in (5) with $2\mu_B$ or $2\mu_F$, we obtain the familiar expressions for chemical potentials of composite bosons consisting of either two bosons [10, 11] or two fermions [12, 13]. The crossover temperature T_* plays the role of a pseudogap temperature, and therefore the Green functions of elementary fermions and bosons acquire a two-pole structure below T_* in similarity with [13].

For lower temperatures $T_0 < T < T_*$, where

$$T_0 = \frac{2\pi n}{m_F + m_B}$$

is the degeneracy temperature of composite fermions, the numbers of elementary fermions and bosons are exponentially small. The chemical potential of composite fermions is given by

$$\mu_{\text{comp}} = -T\ln(T/T_0),$$

and hence

$$|\mu_{\text{comp}}| \ll |E_b| \quad \text{for } T \ll T_*.$$

By the Hubbard–Stratonovich transformation, the original partition function

$$Z = \int \mathcal{D}\bar{b}\mathcal{D}b\mathcal{D}\bar{c}\mathcal{D}c\exp\{-\beta F\}$$

can be written in terms of the composite fermions,

$$Z = \int \mathcal{D}\bar{\Psi}_\alpha \mathcal{D}\Psi_\alpha \exp\{-\beta F_{\text{eff}}\}.$$

This procedure gives the magnitude of the interaction between the composite fermions. The lowest order of the series expansion is given in Fig. 1. Analytically, this diagram is given by

$$-\frac{1}{2}\sum_n \int \frac{d^2p}{(2\pi)^2}\{G_F^2(\mathbf{p}; i\omega_{nF})G_B^2(-\mathbf{p}; -i\omega_{nB}) + G_F^2(-\mathbf{p}; -i\omega_{nF})G_B^2(\mathbf{p}; i\omega_{nB})\}, \tag{8}$$

where

$$G_F = \frac{1}{i\omega_{nF} - \xi(\mathbf{p})},$$

$$G_B = \frac{1}{i\omega_{nB} - \eta(\mathbf{p})}$$

are the fermion and boson Matsubara Green functions, and

$$\omega_{nF} = (2n+1)\pi T, \quad \omega_{nB} = 2n\pi T$$

are the fermion and boson Matsubara frequencies. This integral actually determines the coefficient b at Ψ^4 in the effective action. Evaluation of integral (8) yields

$$b \approx -\frac{N(0)}{|E_b|^2}, \tag{9}$$

where

$$N(0) = m_{BF}/2\pi.$$

The corrections to the coefficient b are presented in Fig. 2. They explicitly contain the T matrices for the

boson–boson and fermion–fermion interactions. In the intermediate coupling case, these diagrams are small in the small parameters

$$f_{BB0} \sim \frac{1}{\ln(W_B/|E_b|)}, \quad f_{FF0} \sim \frac{1}{\ln(W_F/|E_b|)}.$$

Therefore, the exchange diagram indeed makes the main contribution to the coefficient *b*.

The coefficient of the quadratic term Ψ^2 in the effective action, in agreement with general rules of diagrammatic technique [8], is given by

$$a + \frac{cq^2}{2(m_B + m_F)} = -\frac{1}{\Gamma(q; 0)}, \tag{10}$$

where $\Gamma(q; 0)$ is given by (3). The solution of (10) yields

$$c = \frac{N(0)}{|E_b|}, \quad a = N(0)\ln(T/T_*).$$

Therefore, although T_* in reality corresponds to a smooth crossover and not to a real second-order phase transition, the effective action of composite fermions at temperatures $T \sim T_*$ formally resembles the Ginzburg–Landau functional for the Grassmann field Ψ_α.

If we want to rewrite the effective action with gradient terms

$$\Delta F = \alpha\bar{\Psi}_\alpha\Psi_\alpha + \frac{c}{2(m_F + m_B)}(\nabla\bar{\Psi}_\alpha)(\nabla\Psi_\alpha) + \frac{1}{2}b\bar{\Psi}_\alpha\bar{\Psi}_\beta\Psi_\beta\Psi_\alpha \tag{11}$$

in the form of the energy functional of a nonlinear Schrödinger equation for the composite particle with the mass $m_B + m_F$, we have to introduce the effective order parameter

$$\Delta_\alpha = \sqrt{c}\Psi_\alpha.$$

Accordingly, in terms of Δ_α, the new coefficients $\tilde{a}$ and $\tilde{b}$ at the quadratic and quartic terms become

$$\tilde{a} = \frac{a}{c}, \quad \tilde{b} = \frac{b}{c^2}.$$

We note that the Grassmann field Δ_α corresponds to the composite fermions and is normalized by the condition

$$\Delta_\alpha^+\Delta_\alpha = n_{\text{comp}}.$$

Hence, the coefficient $\tilde{b}$ plays the role of the effective interaction between composite particles. From Eqs. (9) and (10),

$$\tilde{b} = -\frac{1}{N(0)}.$$

This result coincides in absolute value with the result in [14], but has the opposite sign. In [14], the residual interaction between two composite bosons, each consisting of two elementary fermions, was calculated in the two-dimensional case. The difference in sign between these two results is due to different statistics of elementary particles in the two cases. It is also important to calculate $b(q)$, where the momenta of the incoming composite fermions are equal to $(\mathbf{q}, -\mathbf{q})$. It is easy to find that

$$b(q) = -\frac{1}{2}\sum_n\int\frac{d^2p}{(2\pi)^2} \times \{G_B(\mathbf{p}; i\omega_{nB})G_F(\mathbf{p}; -i\omega_{nF}) \times G_B(\mathbf{p}+\mathbf{q}; i\omega_{nB})G_F(\mathbf{p}-\mathbf{q}; -i\omega_{nF}) + G_B(\mathbf{p}; -i\omega_{nB})G_F(\mathbf{p}; i\omega_{nF}) \times G_B(\mathbf{p}-\mathbf{q}; -i\omega_{nB})G_F(\mathbf{p}+\mathbf{q}; i\omega_{nF})\}. \tag{12}$$

In the case of equal masses $m_B = m_F = m$, a straightforward calculation for small q yields

$$b(q) = -\frac{m}{4\pi(|E_b| + q^2/4m)^2}. \tag{13}$$

Accordingly,

$$\tilde{b} = \frac{b}{c^2} \approx -\frac{4\pi}{m(1 + q^2/4m|E_b|)^2}, \tag{14}$$

where

$$|E_b| = \frac{1}{ma^2}.$$

A similar result in the three-dimensional case was obtained in [15]. Hence, the four-particle interaction has a Yukawa form in momentum space. Therefore,

$$U_4(r) \approx -\frac{1}{ma^2}\sqrt{\frac{a}{2r}}\exp\left(-\frac{2r}{a}\right)$$

corresponds to an attractive potential with the interaction radius equal to $a/2$. We can now calculate the binding energy $|E_4|$ of quartets. A straightforward calculation, absolutely similar to the calculation of $|E_b|$, yields

$$1 = \frac{|\tilde{b}|(m_B + m_F)}{2\pi}\int_0^{2/a}\frac{qdq}{q^2 + (m_B + m_F)|E_4|}. \tag{15}$$

Hence,

$$|E_4| = \frac{4}{a^2(m_B + m_F)\left[\exp\left(\frac{4\pi}{|\tilde{b}|(m_B + m_F)}\right) - 1\right]}. \tag{16}$$

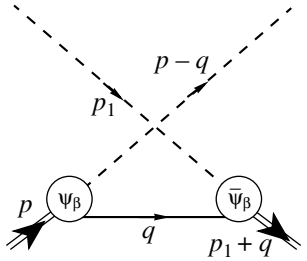

Fig. 3. Exchange diagram for the three-particle interaction.

For equal masses $m_B = m_F$,

$$\frac{|\tilde{b}|(m_B + m_F)}{4} = \frac{1}{2}$$

and therefore

$$|E_4| = \frac{2|E_b|}{(e^{1/2} - 1)} \approx 3|E_b|. \quad (17)$$

The dynamical equilibrium (composite fermion + composite fermion $\rightleftharpoons$ quartet) is again governed by the Saha formula

$$\frac{n_{\text{comp}}^2}{n_4} = \frac{m_4 F}{2\pi}\exp\left\{-\frac{|E_4|}{T}\right\}, \quad (18)$$

where

$$m_4 = \frac{m_B + m_F}{2}.$$

The number of composite fermions is equal to half the number of quartets, $n_4 = n_2/2$, for the crossover temperature

$$T_{**}^{(4)} = \frac{|E_4|}{\ln(|E_4|/2T_0)}. \quad (19)$$

Below this temperature, the quartets of the type $\langle c_{i\uparrow}b_i; c_{j\downarrow}b_j\rangle$ play the dominant role in the system. We note that $T_{**}^{(4)} > T_*$, and therefore quartets dominate over pairs (composite fermions) in the entire temperature interval. We also note that the quartets are in the spin-singlet state. The creation of spin-triplet quartets is prohibited or at least strongly reduced by the Pauli principle. The triplet *p*-wave pairs of composite fermions are possibly created in the strong-coupling case $|E_b| > W$ when the corrections to the coefficient *b* given by the diagrams in Fig. 2 are large and repulsive. However, small parameters are absent in this case, and it is very difficult to control the diagrammatic expansion.

3. THREE-PARTICLE PROBLEM

If we consider the scattering process of an elementary fermion on a composite fermion, we obtain a repulsive sign of the interaction regardless of the relative spin orientation of the composite and elementary fermions. The same result in three dimensions for scattering of an elementary fermion on a dimer consisting of two fermions was obtained in [16]. However, for a scattering process of an elementary boson on a composite fermion, we obtain an attractive sign of the interaction. Moreover, in the two-dimensional case, the Fourier component of the three-particle interaction for $m_B = m_F = m$ is given by (see Fig. 3)

$$U_3(q) = \frac{1}{c}G_F(0, q) = -\frac{8\pi}{m(1 + q^2a^2)}, \quad (20)$$

where $G_F(0, q)$ is the Green function of elementary fermions and $c = N(0)/|E_b|$. Hence,

$$U_3(r) \sim -\frac{1}{ma^2}K_0(r/a) \sim -\frac{1}{ma^2}\sqrt{\frac{a}{r}}\exp(-r/a), \quad (21)$$

which again corresponds to an attractive potential of the Yukawa type, but now with the interaction range equal to *a*. Calculation of the three-particle bound-state energy yields

$$1 = \frac{|U_3(0)|}{2\pi}\int_0^{1/a}\frac{qdq}{q^2/2m_B + q^2/2(m_B + m_F) + |E_3|}. \quad (22)$$

Hence, for $m_B = m_F = m$, we have

$$|E_3| = \frac{3}{4ma^2}\frac{1}{[\exp(3\pi/m|U_3|) - 1]} = \frac{3|E_b|}{4(e^{3/8} - 1)} \approx 1.65|E_b|. \quad (23)$$

We note that we are studying trios and quartets in the zeroth-order exchange approximation. A more rigorous solution of the three- and four-particle problems requires analyzing the Skorniakov–Ter-Martirosian type of equations [17]. This investigation will be the subject of a separate publication. The dynamical equilibrium of the type composite fermion + boson $\rightleftharpoons$ trio is governed by the Saha formula

$$\frac{n_B n_{\text{comp}}}{n_3} = \frac{m_3 T}{2\pi}\exp\left\{-\frac{|E_3|}{T}\right\}, \quad (24)$$

where

$$m_3 = \frac{m_B(m_B + m_F)}{2m_B + m_F}.$$

Accordingly, trios dominate over unbound bosons for temperatures $T < T_{**}^{(3)}$, where

$$T_{**}^{(3)} = \frac{|E_3|}{\ln(|E_3|/2T_0)}. \quad (25)$$

We note that $T_{**}^{(3)} < T_{**}^{(4)}$, and therefore, trios are not as important as quartets.

As a result, there are mostly quartets in the system for $T < T_{**}^{(4)}$. The quartets are Bose-condensed at the critical temperature

$$T_c = \frac{T_0}{8\ln\ln(4/na^2)}$$

in the case of equal masses. It is important to note that in the Feshbach resonance scheme [4, 5, 18], we are usually in the regime $T \sim T_0$, where quartets prevail over trios and pairs. In this scheme, the particles are first cooled to very low temperatures $T < T_0$ and only then is the sign of the scattering length changed by magnetic field to support the formation of bound pairs. We emphasize that in the restricted geometry of magnetic or optical dipole traps, our theory is valid under the condition $T_c > \omega$, where ω is the level spacing in the trap. For a large number of particles $N \gg 1$ in the two-dimensional trap, $\omega \sim T_0/N^{1/2}$ ($\omega \sim T_0/N^{1/3}$ in three-dimensional traps), and this condition is therefore easily satisfied. We also note that octets are not formed in the system because two quartets repel each other due to the Pauli principle, in similarity with the results in [14, 19].

4. CONCLUSIONS

We have considered the appearance and pairing of composite fermions in a Fermi–Bose mixture with an attractive interaction between fermions and bosons.

At equal densities of elementary fermions and bosons, the system is described at low temperatures by a one-component attractive Fermi gas for composite fermions and is unstable with respect to formation of quartets.

The problem that we considered is important for a theoretical understanding of high-temperature superconducting materials and for the investigation of Fermi–Bose mixtures of neutral particles at low and ultralow temperatures. In high-T_c superconductors, quartets play the role of singlet superconductive pairs. The radius of the quartets (the coherence length of the superconducting pair) is governed by the binding energy $|E_4|$ of the quartets. If $|E_4|$ is larger that T_0, the quartets are local: $p_F a < 1$. For

$$T_c = \frac{T_0}{8\ln\ln(4/na^2)},$$

the local quartets are Bose-condensed and the system becomes superconductive. We note that at higher temperatures $T > T_0$, a certain amount of trios is also present in the system in addition to quartets. The role of trios is usually neglected in the standard theories of high-T_c superconductivity.

We also note that we consider the low-density limit $|E_b| \gg T_0$. In the opposite case of higher densities $T_0 \gg |E_b|$, Bose–Einstein condensation of bosons or bibosons [11, 20, 21] occurs earlier than the creation of composite fermions and quartets. Such a state can be distinguished from an ordinary BCS-superconductor by measuring the temperature dependence of the specific heat and normal density.

For Fermi–Bose mixtures, our investigations have enriched the superfluid phase diagram in magnetic and optical dipole traps and are important in connection with recent experiments where weakly bound dimers $^6Li_2$ and $^{40}K_2$, consisting of two elementary fermions, were observed [22, 23]. We note that in an optical dipole trap, it is possible to obtain an attractive scattering length for fermion–boson interaction with the help of the Feshbach resonance [18]. We also note that even in the absence of the Feshbach resonance, it is experimentally possible now to create a Fermi–Bose mixture with attractive interaction between fermions and bosons. For example, in [24, 25], such a mixture of ^{87}Rb (bosons) and ^{40}K (fermions) was experimentally studied. Moreover, the authors of [24, 25] experimentally observed the collapse of the Fermi gas with the sudden disappearance of fermionic ^{40}K atoms when the system enters the degenerate regime. In principle we cannot exclude the fact that this is just a manifestation of $\langle bc; bc\rangle$ quartets being created in the system. We note that in the regime of strong attraction between fermions and bosons, phase separation with the creation of larger clusters or droplets is also possible. We also note that a much slower collapse in the Bose subsystem of ^{87}Rb atoms can possibly be explained by the fact that the number of Rb atoms in the trap is much larger than the number of K atoms, and therefore, after the formation of composite fermions, many residual bosons are still present in the system. A more thorough comparison of our results with an experimental situation will be the subject of a separate publication. Here, we only mention that for the experiments performed in [24, 25], the three-dimensional case is more relevant. In the three-dimensional case, the attractive interaction between composite fermions acquires the form

$$\tilde{b}(q) = -\frac{\pi a_{\text{eff}}}{m_{BF}[1 + q^2/2(m_B + m_F)|E_b|]}, \qquad (26)$$

where

$$|E_b| = \frac{1}{2m_{BF}a^2}$$

is a shallow level of a fermion–boson bound state. We note that in the case of a repulsive interaction between two bosons (each of which consists of two fermions), $a_{\text{eff}} = 2a$ in the mean-field theory in [19], $a_{\text{eff}} = 0.75a$ in the calculations in [15], and $a_{\text{eff}} = 0.6a$ in the calcula-

tions in [16]. The shallow bound state of quartets exists in the three-dimensional case only if

$$a_{\text{eff}} > 2\pi a\left(\frac{m_{BF}}{m_B + m_F}\right)^{3/2}. \quad (27)$$

For $m_B = m_F = m$, we have

$$a_{\text{eff}} > \pi a/4.$$

ACKNOWLEDGMENTS

The authors acknowledge helpful discussions with A. Andreev, Yu. Kagan, L. Keldysh, B. Meierovich, P. Wölfle, G. Khaliullin, A. Chernyshev, I. Fomin, M. Mar'enko, A. Smirnov, P. Arseev, E. Maksimov, G. Shlyapnikov, M. Baranov, and A. Sudbø. This work was supported by the Russian Foundation for Basic Research (project no. 02-02-17520), Russian President Program for Science Support (grant no. 00-15-96-9694), and by the Grant of Russian Academy of Sciences for Young Scientists.

REFERENCES

1. J. Bardeen, G. Baym, and D. Pines, Phys. Rev. **156**, 207 (1967).
2. J. Ranninger and S. Robaszkiewicz, Physica B (Amsterdam) **135**, 468 (1985).
3. B. K. Chakraverty, J. Ranninger, and D. Feinberg, Phys. Rev. Lett. **81**, 433 (1998).
4. W. Hofstetter, J. I. Cirac, P. Zoller, *et al.*, cond-mat/0204237.
5. J. N. Milstein, S. J. J. M. F. Kokkelmans, and M. J. Holland, cond-mat/0204334.
6. D. V. Efremov and L. Viverit, Phys. Rev. B **65**, 134519 (2002).
7. P. Nozieres and S. Schmittt-Rink, J. Low Temp. Phys. **59**, 195 (1985).
8. A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskii, *Methods of Quantum Field Theory in Statistical Physics* (Fizmatgiz, Moscow, 1962; Dover, New York, 1963).
9. L. D. Landau and E. M. Lifshitz, *Statistical Physics*, 4th ed. (Nauka, Moscow, 1995; Butterworth, London, 1999).
10. P. Nozieres and D. Saint James, J. Phys. (Paris) **4**, 1133 (1982).
11. M. Yu. Kagan and D. V. Efremov, Phys. Rev. B **65**, 195103 (2002).
12. M. Yu. Kagan, R. Fresard, M. Capezzali, and H. Beck, Phys. Rev. B **57**, 5995 (1998).
13. M. Yu. Kagan, R. Fresard, M. Capezzali, and H. Beck, Physica B (Amsterdam) **284**–**288**, 447 (2000).
14. M. Drechsler and W. Zwerger, Ann. Phys. (Leipzig) **1**, 15 (1992).
15. P. Pieri and G. C. Strinati, Phys. Rev. B **61**, 15370 (2000).
16. D. S. Petrov, C. Salomon, and G. V. Shlyapnikov, cond-mat/0309010.
17. G. V. Skorniakov and K. A. Ter-Martirosian, Zh. Éksp. Teor. Fiz. **31**, 775 (1956) [Sov. Phys. JETP **4**, 648 (1956)].
18. E. Timmermans, P. Tommasini, M. Hussein, and A. Kerman, Phys. Rep. **315**, 199 (1999).
19. R. Haussmann, Z. Phys. B **91**, 291 (1993).
20. P. A. Lee and N. Nagaosa, Phys. Rev. B **46**, 5621 (1992).
21. P. A. Lee, N. Nagaosa, T. K. Ng, and X. G. Wen, Phys. Rev. B **57**, 6003 (1998).
22. C. A. Regal, C. Ticknor, J. L. Bohn, and D. S. Jin, Nature **424**, 47 (2003).
23. B. G. Levi, Phys. Today **56**, 21 (2003).
24. G. Roati, F. Riboli, G. Modugno, and M. Inguscio, Phys. Rev. Lett. **89**, 150403 (2002).
25. G. Modugno, G. Roati, F. Riboli, *et al.*, Science **297**, 2240 (2002).

Journal of Experimental and Theoretical Physics, Vol. 99, No. 3, 2004, pp. 647–662.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 3, 2004, pp. 741–757.

NONLINEAR PHYSICS

Nonlinear Regimes of Ultrashort Pulse Propagation in an Array of Anisotropic Tunneling States

S. V. Nesterov[a,*] and S. V. Sazonov[b]

[a]*Tomsk State University, Tomsk, 634004 Russia*
[b]*Kaliningrad State University, Kaliningrad, 236041 Russia*
**e-mail: nst@alg.kaliningrad.ru*

Received October 8, 2003

Abstract—Dynamics of an ultrashort electromagnetic pulse in a system with an array of anisotropic tunneling states spanned by the pulse spectrum are analyzed. A system of nonlinear wave equations is derived for the ordinary and extraordinary components of the pulse propagating at an arbitrary angle to the anisotropy axis. Different regimes of ultrashort pulse propagation parallel and perpendicular to the anisotropy axis are examined. Ultrashort-pulse propagation regimes analogous to self-induced transparency and extraordinary transparency are identified. The properties of rational soliton-like pulses having no quasi-monochromatic analogs are analyzed. A longitudinal electric field component is generated in each regime, whereas off-resonance quasi-monochromatic pulses propagating under similar conditions (parallel and perpendicular to the anisotropy axis) have no longitudinal components. Stability of the solutions obtained and the effect of diffraction on ultrashort pulse dynamics are analyzed. The values of pulse parameters for which defocusing dominates over self-focusing are calculated. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Studies of the nonlinear dynamics of few-cycle laser pulses, also called ultrashort pulses (USP), have developed into a well-established line of research [1–18]. Since pulses of this kind cannot be characterized by well-defined carrier frequencies, they cannot be modeled by invoking the slowly varying envelope approximation routinely applied to quasi-monochromatic pulses [19]. For this reason, self-induced transparency (which has been well studied for quasi-monochromatic pulses [20–22]) has certain distinctive characteristics in the case of USP propagation [23, 24]. In [24], this phenomenon was analyzed for a USP propagating in a hydrogen-bonded ferroelectric near the Curie temperature T_c. In such a system, the pulse interacts with tunneling proton states giving rise to a soft mode. Since the soft mode is strongly suppressed by dipole–dipole interactions between tunneling states near T_c, self-induced transparency cannot be observed for resonant quasi-monochromatic pulses. However, a high-energy USP whose spectrum spans the tunneling states can eliminate dipole–dipole interactions [24]. Therefore, tunneling states exhibit individual rather than collective behavior and tunneling modes are neither soft nor suppressed in the region occupied by the pulse. As a result, the pulse propagates in a soliton regime involving substantial excitation of tunneling states, which is analogous to self-induced transparency. Tunneling states in a KDP crystal are highly anisotropic: proton tunneling mainly occurs in the plane perpendicular to the ferroelectric axis [25, 26]. Accordingly, these tunneling states in double-well potentials are treated as one-dimensional in theoretical studies [24–27]. More commonly, tunneling states arise in a three-dimensional array of anisotropic potential barriers separating potential minima. Such potentials are created by the crystalline electric field, which eliminates degeneracy in the absolute value of the projection of orbital angular momentum. Thus, tunneling states are doubly degenerate in the general case. Moreover, anisotropy can give rise to a permanent dipole moment, which corresponds to nonzero diagonal matrix elements of the dipole moment operator. A permanent dipole moment can strongly affect nonlinear regimes of both quasi-monochromatic-pulse and USP propagation in such media [28–32].

In [32], use was made of the fact that an optical pulse consists of two components called ordinary and extraordinary waves. In particular, it was demonstrated that a resonant quasi-monochromatic ordinary wave can generate an ultrashort extraordinary-wave pulse (video pulse), which gives rise to a dynamic shift in quantum-transition frequency due to permanent dipole moment. Subsequently, the two components are coupled via nonlinear interaction, and soliton regimes of two-component pulse propagation become possible. When the ordinary component is dominant, the resonant phenomenon of self-induced transparency is observed. In the opposite limit, the regime called extraordinary transparency in [32] takes place. In the latter case, the two-component solitons decelerate as in the case of self-induced transparency and have no appreciable dynamic effect on the populations of quantum states. In view of the aforementioned difference

between nonlinear regimes of USP and quasi-monochromatic pulse propagation, it would be interesting to examine dynamics of a two-component pulse in the case when neither an ordinary nor extraordinary component has a well-defined carrier frequency; i.e., it must be classified as an ultrashort pulse by standard terminology. For this reason, two-component USP propagation in an array of anisotropic tunneling states is analyzed in the present study.

The paper is organized as follows. In Section 2, we formulate a model and use it as a basis for deriving a system of material and wave equations describing self-consistent dynamics of tunneling states characterized by uniaxial anisotropy and USP propagation in an array of such states. In Section 3, the fact that the USP spectrum spans the tunneling states is used to eliminate material parameters and to derive a system of nonlinear wave equations for the pulse components. Section 4 deals with several regimes of two-component USP propagation parallel and perpendicular to the anisotropy axis. Regimes analogous to self-induced transparency and extraordinary transparency are identified. In Section 5, the properties of rational soliton-like pulses having no quasi-monochromatic analogs are analyzed. In Section 6, we examine the effect of transverse perturbations, such as diffraction, on the dynamics of the one-dimensional solitary pulses considered in the preceding section. In the Conclusions section, we summarize the main results of the study and outline some promising directions of further research.

2. SYSTEM OF MATERIAL AND WAVE EQUATIONS

In the case of a double-well potential, the ground-state and first excited-state wavefunctions of a tunneling particle are, respectively, symmetric and antisymmetric under the coordinate inversion $z' \longrightarrow -z'$ of the tunneling axis [25, 26]. Since the probability densities associated with the corresponding wavefunctions are identical and finite in both wells, tunneling can occur. This particular model is generally used to describe an order–disorder type ferroelectric [25, 26].

A one-dimensional model can be generalized by extending a double-well potential in directions perpendicular to the z' axis. Suppose that both the wells and the potential barrier between them are axially symmetric. The uniaxial anisotropy is attributed to the crystalline electric field. This implies that the energy eigenstates of the tunneling particle are degenerate in the absolute value of the projection m of orbital angular momentum l on the z' axis (henceforth, we neglect spin-related effects). In such a double-well potential, $m = 0$ for the ground state and $m = \pm 1$ for the first excited state.

The operator of interaction between the particle and the internal electric field responsible for the uniaxial anisotropy is invariant under the inversions

$$z' \longrightarrow -z', \quad \varphi \longrightarrow -\varphi,$$

where φ is the angle in a cylindrical coordinate system. Therefore, we can consider even and odd wavefunctions of the particle with respect to these symmetries:

$$\psi(r, z', \varphi) = \pm\psi(r, -z', -\varphi').$$

The ground level (with $m = 0$) is even, and the excited one (with $m = \pm 1$) is odd and doubly degenerate. Note that the concept of parity is defined here with respect to the inversions specified above, rather than to the inversion of the entire coordinate system (as in the case of spherical symmetry). This restricted symmetry implies that an energy eigenstate can have a permanent dipole moment (see below).

The three-level model with a doubly degenerate upper level can be considered as a three-dimensional, axially symmetric generalization of the two-level system corresponding to a one-dimensional double-well potential.

The tunneling frequency ω_0 is on the order of 10^{13} s^{-1}.

Following [18, 32], we represent the energy eigenstates of a proton as

$$\psi_{\mu, m} = R_{\mu, m}(r, z')\exp(im\varphi), \tag{1}$$

where μ is the set of quantum numbers associated with cylindrical symmetry.

Then, we can find the $\mu \longleftrightarrow \nu$ dipole transition matrix elements [18, 32]. The corresponding Cartesian components are

$$d^x_{\mu\nu} = \frac{d_{\mu\nu}}{\sqrt{2}}|\Delta m_{\mu\nu}|, \quad d^{y'}_{\mu\nu} = i\frac{d_{\mu\nu}}{\sqrt{2}}\Delta m_{\mu\nu},$$

$$d^{z'}_{\mu\nu} = D_{\mu\nu}(1 - |\Delta m_{\mu\nu}|),$$

with

$$\Delta m_{\mu\nu} = m_\mu - m_\nu = 0, \pm 1,$$

$$d_{\mu\nu} = -\sqrt{2}\pi e\int_0^\infty r^2 dr \int_{-\infty}^{+\infty} R_\mu(r, z')R_\nu(r, z')dz',$$

$$D_{\mu\nu} = -2\pi e\int_0^\infty r dr \int_{-\infty}^{+\infty} z' R_\mu(r, z')R_\nu(r, z')dz'.$$

Here, x, y', and z' are the axes of a Cartesian orthogonal coordinate system, and e is the charge of a tunneling particle.

By virtue of the wavefunction symmetry indicated above, it holds that

$$R_{\mu,m}(r,z') = \pm R_{\mu,-m}(r,-z'),$$

with plus and minus signs corresponding to $m = 0$ and $m = \pm 1$, respectively. The function $R_{\mu,m}(r,z')$ cannot be characterized by any definite parity under the inversion $z' \longrightarrow -z'$ when m is held invariant. Therefore, the diagonal elements $D_{\mu\mu}$ given by the expression for $D_{\mu\nu}$ have nonzero values in the general case.

To analyze USP propagation at an arbitrary angle α to the anisotropy axis, we perform a rotation about the x axis. In the new Cartesian system (x, y, z), the dipole-moment components are written as

$$d^x_{\mu\nu} = \frac{d_{\mu\nu}}{\sqrt{2}}|\Delta m_{\mu\nu}|,$$

$$d^y_{\mu\nu} = i\frac{d_{\mu\nu}}{\sqrt{2}}\Delta m_{\mu\nu}\cos\alpha - D_{\mu\nu}(1-|\Delta m_{\mu\nu}|)\sin\alpha, \quad (2)$$

$$d^z_{\mu\nu} = i\frac{d_{\mu\nu}}{\sqrt{2}}\Delta m_{\mu\nu}\sin\alpha + D_{\mu\nu}(1-|\Delta m_{\mu\nu}|)\cos\alpha.$$

Accordingly, the matrix elements of the Hamiltonian $\hat{V}$ of electric-field–dipole interaction between a tunneling proton state and the pulse field are expressed as

$$\begin{aligned} V_{\mu\nu} &= -\mathbf{d}_{\mu\nu}\cdot\mathbf{E}(\mathbf{r},t) \\ &= \left[D_{\mu\nu}(1-|\Delta m_{\mu\nu}|)\sin\alpha - i\frac{d_{\mu\nu}}{\sqrt{2}}\Delta m_{\mu\nu}\cos\alpha\right]E_e \\ &\quad - \frac{d_{\mu\nu}}{\sqrt{2}}|\Delta m_{\mu\nu}|E_o \\ &\quad - \left[D_{\mu\nu}(1-|\Delta m_{\mu\nu}|)\cos\alpha + i\frac{d_{\mu\nu}}{\sqrt{2}}\Delta m_{\mu\nu}\sin\alpha\right]E_z. \end{aligned} \quad (3)$$

Here, E_o and E_e denote the electric fields associated with the ordinary and extraordinary components (parallel to the x and y axes, respectively), and E_z is the longitudinal component of the pulse, which is always present (even if weak) in electromagnetic wave propagation in crystals [33].

For the degenerate quantum transition considered here ($\mu = 1$ for $m = 0$ and $\mu = 2, 3$ for $m = \pm 1$), the system of material equations for the elements of the density matrix $\hat{\boldsymbol{\rho}}$ is written as

$$\frac{\partial\rho_{\mu\nu}}{\partial t} = -i\omega_{\mu\nu}\rho_{\mu\nu} + i[\hat{\boldsymbol{\Omega}},\hat{\boldsymbol{\rho}}]_{\mu\nu}, \quad (4)$$

where $\omega_{21} = \omega_{31} = \omega_0$,

$$\hat{\boldsymbol{\Omega}} = \begin{pmatrix} \Omega_{22} & 0 & \Omega_{31} \\ 0 & \Omega_{22} & \Omega_{21} \\ \Omega^*_{31} & \Omega^*_{21} & \Omega_{11} \end{pmatrix},$$

$$\hat{\boldsymbol{\rho}} = \begin{pmatrix} \rho_{33} & \rho_{32} & \rho_{31} \\ \rho^*_{32} & \rho_{22} & \rho_{21} \\ \rho^*_{31} & \rho^*_{21} & \rho_{11} \end{pmatrix}, \quad (5)$$

$$\Omega_{\mu\mu} = \frac{D_{\mu\mu}E_e}{\hbar}\sin\alpha,$$

$$\Omega_{31} = \frac{d_{31}}{\sqrt{2}}(E_o - iE_e\cos\alpha), \quad (6)$$

$$\Omega_{21} = \frac{d_{21}}{\sqrt{21}}(E_o + iE_e\cos\alpha).$$

In system (4)–(6), the longitudinal component of the pulse is neglected because

$$E_z \ll E_o, E_e.$$

This component is taken into account in the wave equations formulated below, where it is attributed to the transverse polarization of the medium. This approach can be interpreted as a method of successive approximations. Moreover, we set $D_{33} = D_{22}$ for simplicity, following [30]. We should make yet another remark concerning system (4)–(6). In this study, we consider tunneling that may occur in different media, including ferroelectrics. Generally, we should allow for dipole–dipole interaction between different tunneling states, because interaction of this particular type is responsible for the phase transition into the ferroelectric state [25, 26]. However, as we noted above, a high-energy pulse whose spectrum spans the tunneling states eliminates the dipole–dipole interactions, and each tunneling state interacts with the pulse field much more strongly than with similar neighboring tunneling states [24]. For this reason, we neglect dipole–dipole interaction here, and all conclusions drawn below apply to ferroelectrics as well.

Equations (4)–(6) are supplemented with the Maxwell equation for the electric field of the pulse:

$$\Delta\mathbf{E} - \nabla(\nabla\mathbf{E}) - \frac{\hat{\mathbf{n}}^2}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2} = \frac{4\pi}{c^2}\frac{\partial^2\mathbf{P}}{\partial t^2}, \quad (7)$$

where c is the speed of light in free space, $\hat{\mathbf{n}}$ is the refractive index tensor due to off-resonant transitions different from the tunneling considered here, and $\mathbf{P}$ is the polarization of the medium.

Hereinafter, we assume that the parameters of a pulse propagating along the z axis weakly depend on the transverse coordinates y and x. Since $E_z \ll E_o, E_e$, the transverse components of the term $\nabla(\nabla\mathbf{E})$ can be neglected. However, the contribution of its z component is important. Retaining only the derivatives with respect to z in this term, integrating the resulting equation with respect to time, and using the fact that both E_z and P_z vanish at infinity, we obtain

$$E_z = -\frac{4\pi P_z}{n_\parallel},$$

where $n_\parallel$ is the longitudinal component of the refractive index.

The expression

$$P_j = N\mathrm{Sp}(\hat{\boldsymbol{\rho}}\hat{\mathbf{d}}^j),$$

for polarization components (N is the concentration of tunneling states; $j = x, y, z$) yields

$$P_x = P_o = N\left[\frac{d_{21}}{\sqrt{2}}(\rho_{21} + \rho_{21}^*) + \frac{d_{31}}{\sqrt{2}}(\rho_{31} + \rho_{31}^*)\right],$$

$$P_y = P_e = N\left[i\frac{d_{21}}{\sqrt{2}}\cos\alpha(\rho_{21}^* - \rho_{21}) - i\frac{d_{31}}{\sqrt{2}}\cos\alpha(\rho_{31}^* - \rho_{31}) + D\sin\alpha(\rho_{11} - W_1)\right],$$

$$P_z = N\left[\frac{i}{\sqrt{2}}\sin\alpha(d_{31}(\rho_{31}^* - \rho_{31}) - d_{21}(\rho_{21}^* - \rho_{21})) + D\cos\alpha(\rho_{11} - W_1)\right],$$

where $D = D_{11} - D_{22}$. The quantity D is henceforth referred to as the permanent dipole moment. Then, the wave equations for E_o and E_e become

$$\Delta E_o - \frac{n_o^2}{c^2}\frac{\partial^2 E_o}{\partial t^2} = \frac{4\pi N}{c^2\sqrt{2}} \times \frac{\partial^2}{\partial t^2}[d_{21}(\rho_{21} + \rho_{21}^*) + d_{31}(\rho_{31} + \rho_{31}^*)], \quad (8)$$

$$\Delta E_e - \frac{n_e^2}{c^2}\frac{\partial^2 E_e}{\partial t^2} = \frac{4\pi N}{c^2}\frac{\partial^2}{\partial t^2} \times \left\{i\frac{\cos\alpha}{\sqrt{2}}[d_{21}(\rho_{21}^* - \rho_{21}) - d_{31}(\rho_{31}^* - \rho_{31})] + D\sin\alpha(\rho_{11} - W_1)\right\}, \quad (9)$$

where n_o and n_e are the refractive indices corresponding to the ordinary and extraordinary waves, respectively. The longitudinal field component is

$$E_z = -\frac{4\pi N}{n_\parallel}\left\{i\frac{\sin\alpha}{\sqrt{2}}[d_{31}(\rho_{31}^* - \rho_{31}) - d_{21}(\rho_{21}^* - \rho_{21})] + D\cos\alpha(\rho_{11} - W_1)\right\}, \quad (10)$$

where W_j ($j = 1, 2$) denotes the initial level populations ($W_3 = W_2$ since the upper level is degenerate).

System (4)–(6), (8)–(10) provides a self-consistent description of the pulse components and the array of tunneling states. According to these equations, the ordinary wave causes transitions in the system of tunneling sublevels, whereas the extraordinary component both induces these transitions and gives rise to a dynamic shift in the transition frequency due to a permanent dipole moment (nonzero diagonal elements of the matrix $\hat{\boldsymbol{\Omega}}$). According to (10), the longitudinal component E_z is generated by a permanent dipole moment and transition dipole moments.

When $\alpha = 0$, only transitions are caused by both ordinary and extraordinary components. When $\alpha = 90°$, the effects of these components are different: whereas the ordinary one causes quantum transitions, the extraordinary one only shifts the transition frequency. The longitudinal component is determined by the permanent dipole moment in the former case and by the transition dipole moments in the latter. Note that neither off-resonance pulses nor continuous radiation have longitudinal components when $\alpha = 0$ or 90° [18, 33]. According to (10) and the analysis presented below, this is not true for USP propagation.

In deriving a self-consistent system of material and wave equations, we assumed that the pulse field interacts only with tunneling states at the frequency ω_0. Since systems with double-well potentials generally have higher levels (proton states), the model may not be reducible to a two-level system with a degenerate upper level. However, the nearest higher level in many cases

is separated by a considerable gap: while $\omega_0 \sim 10^{13}$ s^{-1}, the frequency-scale distance to such a level is about 5×10^{14}–10^{15} s^{-1} in a KDP-type ferroelectric [26]. Since the spectrum of a pulse with duration $\tau_p \sim 10^{-14}$ s does not extend to the high-lying levels, their influence, as well as that of electronic transitions (for a similar reason), is taken into account by introducing the refractive index tensor.

3. NONLINEAR WAVE EQUATIONS

Now, we use Eqs. (4)–(6) to express the matrix elements of $\hat{\rho}$ in terms of E_o and E_e and proceed to an analysis of nonlinear wave equations. A condition for the tunneling states to be spanned by the USP spectrum can be written as follows [1–3, 7–9]:

$$\omega_0 \tau_p \ll 1. \tag{11}$$

By condition (11), the first term in (4) can be neglected. The resulting equation is written in symbolic form as

$$\frac{\partial \hat{\rho}}{\partial t} = i[\hat{\Omega}, \hat{\rho}]. \tag{12}$$

The formal solution to (12) is

$$\hat{\rho} = \hat{U}\hat{\rho}_0\hat{U}^+, \tag{13}$$

where $\hat{\rho}$ is the density matrix before the interaction with the pulse and $\hat{U}$ is a unitary evolution operator. Expression (5) for $\hat{\Omega}$ implies that this matrix does not commute with an analogous matrix taken at a different time. Following [32, 34–36], we henceforth assume that the pulse is so short and intense that the noncommutativity is not significant within the interval Δt of interaction with the pulse. Accordingly, the evolution operator can approximately be written as follows [32, 34–36]:

$$\hat{U} = \lim_{\substack{\Delta t \to 0 \\ \|\hat{\Omega}\| \to \infty}} [\exp(i\hat{\theta})], \quad \hat{\theta} = \int_{t_0}^{t_0 + \Delta t} \hat{\Omega} dt', \tag{14}$$

where $\|\ldots\|$ is an operator norm.

The approach based on (13) and (14) is equivalent to an operator version of the asymptotic WKB method (see [32, 35–38]). We use Sylvester's formula [39] to calculate the operator exponential in (14):

$$\exp(i\hat{\theta}) = \sum_j \exp(i\lambda_j) \prod_{k \neq j} \frac{\hat{\theta} - \lambda_k \hat{I}}{\lambda_j - \lambda_k}, \tag{15}$$

where $\hat{I}$ is the identity matrix, $\{\lambda_k\}$ is the spectrum of the operator $\hat{\theta}$. Since $\|\hat{\Omega}\| \longrightarrow \infty$, the eigenvalues λ_k tend to a similar limit. Using L'Hôpital's rule to evaluate the indeterminate forms that multiply the imaginary exponentials in (15) and using the relation

$$\lambda_j \approx \int_{t_0}^{t_0 + \Delta t} p_j dt',$$

where $\{p_j\}$ is the spectrum of the operator $\hat{\Omega}$, combine (14) with (15) to obtain

$$\hat{U} = \sum_j \exp\left\{ i \int_{-\infty}^{t} p_j dt' \right\} \prod_{k \neq j} \frac{\hat{\Omega} - p_k \hat{I}}{p_j - p_k}. \tag{16}$$

The eigenvalues of $\hat{\Omega}$ are

$$p_1 = \Omega_{22}, \quad p_{2,3} = (\Omega_{11} + \Omega_{22} \pm |\Omega|)/2,$$

where

$$\begin{aligned} |\Omega| &= \sqrt{\Omega_e^2 + |\Omega_{oe}|^2}, \\ \tilde{\Omega}_e &= \Omega_{11} - \Omega_{22} = \frac{DE_e}{\hbar} \sin\alpha, \\ |\Omega_{oe}|^2 &= 4(|\Omega_{21}|^2 + |\Omega_{31}|^2). \end{aligned} \tag{17}$$

Using (16) and (17), we represent the evolution operator as follows:

$$\begin{aligned} \hat{U} = 1 - 4\frac{(\hat{\Omega} - \Omega_{11}\hat{I})(\hat{\Omega} - \Omega_{22}\hat{I})}{|\Omega_{oe}|^2} \\ \times \left(1 - \exp\left(i\frac{\theta_e}{2}\right)\cos\frac{\theta}{2}\right) + 2i\exp\left(i\frac{\theta_e}{2}\right)\frac{\hat{\Omega} - \Omega_{22}\hat{I}}{\Omega} \\ \times \left(1 - 2\frac{\tilde{\Omega}_e(\hat{\Omega} - \Omega_{11}\hat{I})}{|\Omega_{oe}|^2}\right)\sin\frac{\theta}{2}, \end{aligned} \tag{18}$$

where

$$\theta = \int_{-\infty}^{t} |\Omega| dt', \quad \theta_e = \int_{-\infty}^{t} \tilde{\Omega}_e dt'.$$

By virtue of (13), this yields

$$\rho_{21} = i(W_1 - W_2)\frac{\Omega_{21}}{|\Omega|}\sin\theta + 2(W_1 - W_2)\frac{\Omega_{11}\Omega_{21}}{|\Omega|^2}\sin^2\frac{\theta}{2}, \tag{19}$$

$$\rho_{31} = i(W_1 - W_2)\frac{\Omega_{31}}{|\Omega|}\sin\theta + 2(W_1 - W_2)\frac{\Omega_{11}\Omega_{31}}{|\Omega|^2}\sin^2\frac{\theta}{2}, \tag{20}$$

$$\rho_{32} = -\Omega_{31}\Omega_{21}^* B, \tag{21}$$

$$\rho_{11} = W_1 - (W_1 - W_2)\frac{|\Omega_{oe}|^2}{|\Omega|^2}\sin^2\frac{\theta}{2}, \tag{22}$$

$$\rho_{22} = W_2 + 4(W_1 - W_2)\frac{|\Omega_{21}|^2}{|\Omega|^2}\sin^2\frac{\theta}{2}, \tag{23}$$

$$\rho_{33} = W_2 + 4(W_1 - W_2)\frac{|\Omega_{31}|^2}{|\Omega|^2}\sin^2\frac{\theta}{2}, \tag{24}$$

where B is a real quantity depending on the matrix $\hat{\mathbf{\Omega}}$ (its particular form is not essential for the present analysis).

Next, we approximately take into account the first term on the right-hand side of (4), using expressions (19)–(24). Expressing the elements of $\hat{\mathbf{\Omega}}$ in terms of the original field and material variables, we obtain

$$\frac{\partial}{\partial t}[d_{21}(\rho_{21} + \rho_{21}^*) + d_{31}(\rho_{31} + \rho_{31}^*)] = \frac{2\sqrt{2}}{\hbar}d^2(W_1 - W_2) \times \left[\left(\omega_0 + \frac{DE_e}{\hbar}\sin\alpha\right)\frac{E_o}{|\Omega|}\sin\theta - \delta\cos\alpha E_e\left(\cos\theta - \frac{2\omega_0 D}{\hbar}\sin\alpha\frac{E_e}{|\Omega|^2}\sin^2\frac{\theta}{2}\right)\right], \tag{25}$$

$$\frac{\partial}{\partial t}[d_{21}(\rho_{21} + \rho_{21}^*) - d_{31}(\rho_{31} + \rho_{31}^*)] = -i\frac{2\sqrt{2}}{\hbar}d^2(W_1 - W_2) \times \left[\left(\omega_0 + \frac{DE_e}{\hbar}\sin\alpha\right)\cos\alpha\frac{E_o}{|\Omega|}\sin\theta + \delta\left(\cos\theta - \frac{2\omega_0 D}{\hbar}\sin\alpha\frac{E_e}{|\Omega|^2}\sin^2\frac{\theta}{2}\right)\right], \tag{26}$$

where

$$d^2 = (d_{21}^2 + d_{31}^2)/2, \quad \delta = (d_{21}^2 - d_{31}^2)/2d^2,$$
$$|\Omega| = \frac{1}{\hbar}\sqrt{D^2E_e^2\sin^2\alpha + 4d^2(E_o^2 + E_e^2\cos^2\alpha)}. \tag{27}$$

Substituting (25), (26), and (22) into the right-hand sides of (8) and (9), we invoke the quasi-unidirectional approximation [2–4, 7, 18, 40] to obtain

$$\frac{\partial E_o}{\partial z} = -g\left\{\left(\omega_0 + \frac{DE_e}{\hbar}\sin\alpha\right)\frac{E_o}{\Omega}\sin\theta - \delta\cos\alpha\left(\cos\theta - \frac{2\omega_0 D}{\hbar}\sin\alpha\frac{E_e}{\Omega^2}\sin^2\frac{\theta}{2}\right)\right\} + \frac{c}{2n_o}\Delta_\perp\int_{-\infty}^{\tau} E_o d\tau', \tag{28}$$

$$\frac{\partial E_e}{\partial z} + \frac{n_e - n_o}{c}\frac{\partial E_e}{\partial \tau} = -g\left\{\cos^2\alpha\left(\omega_0 + \frac{DE_e}{\hbar}\sin\alpha\right)\frac{E_o}{\Omega}\sin\theta - \delta\cos\alpha\left(\cos\theta - \frac{2\omega_0 DE_e}{\hbar\Omega^2}\sin^2\frac{\theta}{2}\right)E_o\right\} + g\left\{\frac{D}{\hbar}\sin\alpha\frac{E_o^2 + E_e^2\cos^2\alpha}{\Omega}\sin\theta\right\} + \frac{c}{2n_o}\Delta_\perp\int_{-\infty}^{\tau} E_e d\tau', \tag{29}$$

where

$$\tau = t - n_o z/c,$$
$$g = \frac{4\pi N}{cn_o\hbar}d^2(W_1 - W_2),$$

and $\Delta_\perp$ is the transverse Laplace operator.

The approximation frequently used in this study is called quasi-unidirectional, because we allow for a weak dependence on the transverse coordinates by retaining the transverse Laplacian in Eqs. (28) and (29). This approach is analogous to the paraxial approximation in the theory of monochromatic beams [41]. Another reason for invoking this approximation calls for a more detailed discussion. First of all, we note that the use of a similar approximation in [40] was based on the assumption of low concentration of atoms interacting with the field,

$$\eta \equiv \frac{4\pi d^2 N}{\hbar\omega_0} \ll 1,$$

which is not required in the present study. Indeed, condition (11) combined with (25) and (26) implies that the polarization associated with both pulse components can be estimated as

$$P_{o,e} \sim \frac{\eta}{4\pi}\omega_0\tau_p\frac{\omega_0 + \tilde{\Omega}_e}{|\Omega|}E_{o,e}\sin\theta.$$

By virtue of (11), the product $\eta\omega_0\tau_p$ can be small even if $\eta \sim 1$. Moreover, since

$$\left|\omega_0 + \tilde{\Omega}_e\right|/|\Omega| < 1,$$

$P_{o,e} \ll E_{o,e}$; i.e., the right-hand side of Eq. (7) is small. The last relation justifies the use of a quasi-unidirectional approximation here.

Combining (10), (19), (20), and (22), we find an expression for the longitudinal field component:

$$E_z = \frac{8\pi N(W_1 - W_2)}{\hbar n_\parallel \Omega}d^2\left\{\delta E_o\sin\alpha\sin\theta + \frac{2D}{\hbar\Omega}\cos\alpha(E_e^2\sin\alpha + E_o^2 + E_e^2\cos^2\alpha)\sin^2\frac{\theta}{2}\right\}. \quad (30)$$

Setting $\alpha = 0$ and 90° in (30), we obtain

$$E_z = 4\pi DN(W_1 - W_2)\sin^2\frac{\theta}{2}, \quad \alpha = 0, \quad (31)$$

$$E_z = \frac{8\pi\delta d^2 N}{n_\parallel}\frac{E_o}{\sqrt{D^2E_e^2 + 4d^2E_o^2}}\sin\theta, \quad \alpha = 90°. \quad (32)$$

Let us consider expressions (31) and (32) in some detail. It is well known that no longitudinal component is generated in a uniaxial crystal when a monochromatic wave propagates parallel or perpendicular to the optical axis [33]. This is also true for USPs that do not contain any resonant Fourier components [18]. The longitudinal component arising in the present model is due to a permanent dipole moment when $\alpha = 0$ and to a shift in the $1 \longleftrightarrow 3$ and $1 \longleftrightarrow 2$ transition dipole moments when $\alpha = 90°$. Furthermore, Eqs. (31) and (32) demonstrate a nonlinear dependence of the longitudinal component on the remaining two field components. Thus, the generation of E_z is an essentially nonlinear phenomenon.

Since the difference between the ordinary and extraordinary waves is not manifested when $\alpha = 0$, either wave can generate a longitudinal component. However, E_z is generated by the ordinary wave in the case of USP propagation perpendicular to the optical axis.

System (28), (29) describes the nonlinear dynamics of the ordinary and extraordinary components of an ultrashort pulse propagating in an array of tunneling states at an arbitrary angle α to the anisotropy axis. The two cases analyzed in detail below correspond to propagation parallel and perpendicular to the anisotropy axis.

4. SELF-INDUCED TRANSPARENCY

If $\alpha = 0$, then $n_e = n_o$, and Eqs. (28) and (29) can be rewritten as the single equation

$$\frac{\partial\Omega}{\partial z} = -\frac{4\pi d^2 N}{n_o\hbar c}(W_1 - W_2) \times\left[\omega_0\frac{\Omega}{|\Omega|}\sin\theta + i\delta\cos\theta\right] + \frac{c}{2n_o}\Delta_\perp\int_{-\infty}^{\tau}\Omega d\tau' \quad (33)$$

for the complex function

$$\Omega = \Omega_o + i\Omega_e = 2d(E_o + iE_e)/\hbar.$$

Representing Ω as

$$\Omega = |\Omega|e^{i\varphi}, \quad (34)$$

we separate the real and imaginary parts of Eq. (33) to obtain

$$\frac{\partial^2\theta}{\partial z\partial\tau} = -\frac{\omega_0}{l}\sin\theta + \frac{c}{2n_o}\Delta_\perp\theta, \quad (35)$$

$$\frac{\partial\varphi}{\partial z} = -\frac{\delta}{l}\cos\theta,$$

where

$$1/l = 4\pi d^2 N(W_1 - W_2)/n_o \hbar c.$$

In deriving Eq. (35), we used the fact that transverse perturbations are small, as well as the fact that the variation of the parameters of the pulse within an interval on the order of τ_p is relatively slow (see above). Accordingly, we can write

$$\exp(-i\varphi(\mathbf{r}, \tau'))\Delta_\perp \int_{-\infty}^{\tau} |\Omega| \exp(i\varphi(\mathbf{r}, \tau'))d\tau'$$
$$\approx \Delta_\perp \int_{-\infty}^{\tau} |\Omega| d\tau' = \Delta_\perp \theta.$$

In the one-dimensional approximation ($\Delta_\perp = 0$), Eq. (35) reduces to the sine-Gordon equation, whose one-soliton solution expressed in terms of z and t is

$$\theta = 4\arctan\left[\exp\left(\frac{t - z/V}{\tau_p}\right)\right], \quad (36)$$

where the velocity is defined by the relation

$$\frac{1}{V} = \frac{n_o}{c} + \frac{\omega_0}{l}\tau_p^2. \quad (37)$$

Using (36) and (31), we find the corresponding expressions for the field components $E_\perp = \hbar|\Omega|/2d$ and E_z and a nonlinear phase φ:

$$E_\perp = \frac{\hbar}{d\tau_p}\operatorname{sech}\left(\frac{t - z/V}{\tau_p}\right), \quad (38)$$

$$E_z = 4\pi DN(W_1 - W_2)\operatorname{sech}^2\left(\frac{t - z/V}{\tau_p}\right), \quad (39)$$

$$\varphi = \frac{\delta}{\omega_0 \tau_p}\left[\frac{t - z/V}{\tau_p} - 2\tanh\left(\frac{t - z/V}{\tau_p}\right)\right]. \quad (40)$$

The tunneling-state populations are determined from (36), (22)–(24), (27):

$$\rho_{11} = W_1 + (W_1 - W_2)\operatorname{sech}^2\left(\frac{t - z/V}{\tau_p}\right),$$

$$\rho_{22} = W_2 - (W_1 - W_2)\frac{d_{21}^2}{d_{21}^2 + d_{31}^2}\operatorname{sech}^2\left(\frac{t - z/V}{\tau_p}\right), \quad (41)$$

$$\rho_{33} = W_2 + (W_1 - W_2)\frac{d_{31}^2}{d_{21}^2 + d_{31}^2}\operatorname{sech}^2\left(\frac{t - z/V}{\tau_p}\right).$$

Expressions (41) demonstrate that a propagating soliton causes total inversion in the tunneling system followed by return to the initial state. This means that the USP propagates in the regime of self-induced transparency.

According to (34), (38), and (40), the plane of USP polarization has a twist because $\delta \neq 0$ ($d_{21} \neq d_{31}$). Note that a theory of self-induced transparency for two-frequency quasi-monochromatic pulse propagation in three-level resonant media (including the degenerate case) was developed in [42, 43]. One important distinction of the case considered here from the results of those studies is that the USP has a longitudinal component due to a permanent dipole moment. In view of (38) and (39), the condition $E_z \ll E_\perp$ used at the outset can be rewritten as

$$\omega_0 \tau_p \ll \frac{d}{D\eta}.$$

This condition should be satisfied in addition to (11). When the concentration of tunneling centers is low ($\eta \ll 1$), this inequality holds for a wide range of d/D. Otherwise, the transition dipole moments must dominate over the permanent dipole moment.

Now, suppose that $\alpha = 90°$. In this case, $\tilde{\Omega}_e = \Omega_e$, and Eqs. (28) and (29) become

$$\frac{\partial \Omega_o}{\partial z} = -\beta_o(\omega_0 + \Omega_e)\frac{\Omega_o}{|\Omega|}\sin\theta + \frac{c}{2n_o}\Delta_\perp \int_{-\infty}^{\tau} \Omega_o d\tau', \quad (42)$$

$$\frac{\partial \Omega_e}{\partial z} + \frac{n_e - n_o}{c}\frac{\partial \Omega_e}{\partial \tau} = \beta_e \frac{\Omega_o^2}{|\Omega|}\sin\theta + \frac{c}{2n_o}\Delta_\perp \int_{-\infty}^{\tau} \Omega_e d\tau', \quad (43)$$

where

$$\beta_o = \frac{4\pi d^2 N}{n_o \hbar c}(W_1 - W_2), \quad \beta_e = \frac{4\pi D^2 N}{n_o \hbar c}(W_1 - W_2),$$

$$|\Omega| = \sqrt{\Omega_o^2 + \tilde{\Omega}_e^2}.$$

Neglecting the distinction between n_e and n_o, setting $\Delta_\perp = 0$, and using the fact that the field vanishes at infinity, we derive from (42) and (43) a relation between Ω_o

and Ω_e analogous to that found in [44, 45], where a different physical problem was solved by applying the inverse scattering method of soliton theory:

$$\Omega_o^2 = -\frac{\beta_o}{\beta_e}(\Omega_e^2 + 2\omega_0\Omega_e). \quad (44)$$

In particular, this entails

$$\Omega_e < 0, \quad |\Omega_e| < 2\omega_0.$$

The approximation adopted in this study implies that $\Omega_o \gg |\Omega_e|$, and we can set

$$|\Omega| \approx \Omega_o, \quad \theta \approx \int_{-\infty}^{t} \Omega_o dt'$$

in (42). On the other hand, Eq. (44) (with $\Omega_e = 0$ when $\beta_e = 0$) yields

$$\Omega_e + \omega_0 = \sqrt{\omega_0^2 - (\beta_e/\beta_o)\Omega_o^2}.$$

Then, Eq. (42) becomes

$$\frac{\partial^2\theta}{\partial z\partial\tau} = -\beta_o\sqrt{\omega_0^2 - \frac{\beta_e}{\beta_o}\left(\frac{\partial\theta}{\partial\tau}\right)^2}\sin\theta. \quad (45)$$

A soliton-like solution to Eq. (45) can be written as

$$E_o = \frac{d\varepsilon}{\hbar\tau_p}\frac{\cosh\xi}{1+\varepsilon^2\sinh^2\xi}, \quad (46)$$

where

$$\varepsilon = \sqrt{1-(D/2d)^2(\omega_0\tau_p)^{-2}}, \quad \xi = (t - z/v)/\tau_p.$$

Again, the relation velocity and duration of the pulse are related by (37). The extraordinary component is expressed as

$$E_e = -\frac{\hbar}{D}\omega_0 \times\left[1 - \frac{\sqrt{1+\varepsilon^2[\varepsilon^2(1+\cosh^2\xi)-2](1+\cosh^2\xi)}}{1+\varepsilon^2\sinh^2\xi}\right]. \quad (47)$$

Assuming that

$$4d^2E_o^2 \gg D^2E_e^2$$

in (32), we obtain

$$E_z = \frac{8\pi\delta N}{n_\parallel}\varepsilon\frac{\sinh\xi}{1+\varepsilon^2\sinh^2\xi}. \quad (48)$$

In contrast to the positive and negative humps in the ordinary and extraordinary components, respectively, the longitudinal one has the form of a bipolar solitary pulse.

Using (22)–(24) and (46) under the condition $|\Omega_{oe}|^2 = |\Omega|^2$, we obtain

$$\rho_{11} = W_1 - \frac{W_1 - W_2}{1+\varepsilon^2\sinh^2\xi},$$
$$\rho_{22} = W_2 + 2\frac{d_{21}^2}{d^2}\frac{W_1 - W_2}{1+\varepsilon^2\sinh^2\xi}, \quad (49)$$
$$\rho_{33} = W_2 + 2\frac{d_{31}^2}{d^2}\frac{W_1 - W_2}{1+\varepsilon^2\sinh^2\xi}.$$

According to (46) and (49), a propagating soliton causes total inversion of population followed by return to the initial state; i.e., the regime of self-induced transparency is predicted, as in the case of $\alpha = 0$.

Setting $D = 0$ ($\varepsilon = 1$) in (46), we obtain expression (38) for a soliton described by the sine-Gordon equation. The pulse described by (46) is characterized by a lower amplitude and weaker spatiotemporal localization as compared to the soliton described by the sine-Gordon equation. The expression for ε entails a lower limit for the pulse duration,

$$\omega_0\tau_p > \left(\frac{D}{2d}\right)^2, \quad (50)$$

which must hold in addition to (11).

Both conditions are easily satisfied when $d/D \geq 5$.

5. RATIONAL ULTRASHORT PULSES

Suppose that $\alpha = 90°$ and $n_e \neq n_o$. Furthermore, assume that $\Omega_e \gg \omega_0$ and Ω_o has an arbitrary value irrespective of Ω_e. Therefore, it again holds that $\|\hat{\mathbf{\Omega}}\| \gg \omega_0$. The field components can also have arbitrary values. We can approximately set

$$\Omega_e + \omega_0 \approx \Omega_e$$

in (42). We also assume proportionality of Ω_e to Ω_o:

$$\Omega_e = q\Omega_o, \quad (51)$$

where q is a free parameter.

In this case, Eqs. (42) and (43) have at least one time-independent traveling-wave solution. An analogous case was considered in [46].

The approach used below not only leads to traveling-wave USP solutions, but also makes it possible to analyze their stability.

Using (51) and the assumptions introduced above, we rewrite Eqs. (42) and (43) with $\Delta_\perp = 0$ as follows:

$$\frac{\partial^2\theta}{\partial z\partial\tau} = -\frac{q\beta_o}{\sqrt{1+q^2}}\frac{\partial\theta}{\partial\tau}\sin\theta,$$
$$\frac{\partial^2\theta}{\partial z\partial\tau} + \frac{n_e - n_o}{c}\frac{\partial^2\theta}{\partial\tau^2} = \frac{\beta_e}{q\sqrt{1+q^2}}\frac{\partial\theta}{\partial\tau}\sin\theta. \quad (52)$$

Since Eq. (52) contains the parameter q of the expected solution, this system can be treated as equations only tentatively. Moreover, these equations must be identical when written for the same quantity. A condition for their compatibility can be written after integrating with respect to τ. Then, we have two ansatzes:

$$\frac{\partial\theta}{\partial z} = -2\beta_o\frac{q}{\sqrt{1+q^2}}\sin^2\frac{\theta}{2},$$
$$\frac{\partial\theta}{\partial\tau} = \frac{2c}{q\sqrt{1+q^2}(n_e - n_o)}(\beta_e + \beta_o q^2)\sin^2\frac{\theta}{2}. \quad (53)$$

These ansatzes, as well as (52), are used in the stability analysis presented below.

Seeking a solution to (53) in the form of a solitary traveling wave, we obtain

$$\theta = -2\operatorname{arccot}\xi, \quad (54)$$

where

$$\xi = (t - z/v)/\tau_p,$$

and the velocity v and duration τ_p are related to q as follows:

$$\frac{1}{v} = \frac{1}{c}\left[n_o + (n_e - n_o)\frac{\beta_o q^2}{\beta_e + \beta_o q^2}\right],$$
$$\tau_p = \frac{(n_e - n_o)q\sqrt{1+q^2}}{2c(\beta_e + \beta_o q^2)}. \quad (55)$$

The field and material components are

$$E_o = \frac{\hbar}{d\tau_p\sqrt{1+q^2}}\frac{1}{1+\xi^2},$$
$$E_e = \frac{2\hbar q}{D\tau_p\sqrt{1+q^2}}\frac{1}{1+\xi^2},$$
$$E_z = -\frac{8\pi\delta dN}{n_\parallel\sqrt{1+q^2}}\frac{\xi}{1+\xi^2}, \quad (56)$$

$$\rho_{11} = W_1 - \frac{W_1 - W_2}{1+q^2}\frac{1}{1+\xi^2},$$
$$\rho_{22} = W_2 + 2\frac{d_{21}^2}{d^2}\frac{W_1 - W_2}{1+q^2}\frac{1}{1+\xi^2},$$
$$\rho_{33} = W_2 + 2\frac{d_{31}^2}{d^2}\frac{W_1 - W_2}{1+q^2}\frac{1}{1+\xi^2}. \quad (57)$$

According to (56), the components of field strength decrease toward infinity as powers rather than exponentials. As in the last case considered in the preceding section, both transverse components are unipolar, while the longitudinal component is bipolar. The latter expression in (55) implies that pulses of this kind cannot propagate when $n_o = n_e$. According to the former expression in (55), the USP velocity lies in the interval between c/n_o and c/n_e. When the ordinary component is dominant ($q^2 \ll 1$), we have $v = c/n_o$, the pulse duration is

$$\tau_p \approx \frac{(n_e - n_o)q}{2c\beta_e},$$

and (57) implies that a propagating soliton causes total inversion followed by return to the initial state (regime of self-induced transparency). In the opposite limit, when $v = c/n_e$, the pulse duration is

$$\tau_p \approx \tau_{pe} \equiv \frac{n_e - n_o}{2c\beta_o}$$

and the populations of tunneling sublevels are virtually constant. This regime can be interpreted as analogous to the extraordinary transparency considered in [32]. It differs from extraordinary transparency in that the velocity of a two-component USP does not decrease, remaining almost equal to c/n_e, while τ_p tends to a constant value τ_{pe}. In the case of positive birefringence ($n_e > n_o$), it holds that

$$c/n_e < v < c/n_o.$$

In the case of negative birefringence ($n_e < n_o$),

$$c/n_o < v < c/n_e.$$

Note that the USP duration is the absolute value of τ_p. The pulse duration has a maximum: if $2D^2 > d^2$, then

$$\tau_{p\max} = \tau_m;$$

if $2D^2 < d^2$, then

$$\tau_{p\max} = \tau_m[2(1-2D^2/d^2)]^{-1/2},$$

$$\tau_m = |n_e - n_o|/2c\beta_o.$$

Note also that the rational solitons described by (56) are different from the one-component (scalar) solitons obtained in [31].

To examine the one-dimensional stability of rational USP, we let τ go to infinity in the first ansatz in (53). Then,

$$\theta = A = \int_{-\infty}^{\infty} \Omega d\tau'$$

and (53) becomes

$$\frac{dA}{dz} = -2\beta_o \frac{q}{\sqrt{1+q^2}} \sin^2\frac{A}{2}. \quad (58)$$

An analysis of this equation shows that the integrated "area" A of the incoming USP approaches $2\pi k$ ($k = 0, 1, 2, \ldots$) as it propagates through the medium. A 2π-pulse can evolve if the input satisfies the condition $2\pi < A_0 < 4\pi$, a 4π-pulse develops if $4\pi < A_0 < 6\pi$, and so on. If $2\pi < A_0$, then $A \longrightarrow 0$ as $z \longrightarrow \infty$ and a rational soliton cannot develop. It is clear from (54) that the integrated areas of the solitons are 2π.

The analysis based on ansatz (58) demonstrates the one-sided stability of a USP in the one-dimensional case. Note that the rational soliton develops over the distance $l_{eff} = 2\beta_o$ when the extraordinary component is dominant ($q^2 \gg 1$), which is much shorter than the corresponding distance in the case of a dominant ordinary component.

6. DIFFRACTION EFFECTS

To analyze the effect of transverse perturbation (including diffraction) on the one-dimensional USP propagation considered above, we use the averaged-Lagrangian method [47].

First, we consider the case of self-induced transparency for a USP propagating along the anisotropy axis. Equation (35) can be derived from the Lagrangian density

$$L = \frac{1}{2}\frac{\partial\theta}{\partial z}\frac{\partial\theta}{\partial\tau} - \frac{\omega_0}{l}(1-\cos\theta) - \frac{c}{4n_o}(\nabla_\perp\theta)^2. \quad (59)$$

We seek a trial solution to (35) with $\Delta_\perp \neq 0$ corresponding to (36) in the form

$$\theta = 4\operatorname{arccot}\{\exp[\rho(\tau-\Phi)]\}, \quad (60)$$

where $\rho = \rho(z, \mathbf{r}_\perp)$ and $\Phi = \Phi(z, \mathbf{r}_\perp)$ denote, respectively, slow and fast functions of coordinates to be determined. Substituting (60) into (59) and integrating the result with respect to τ, we obtain the "averaged Lagrangian"

$$\Lambda = \frac{1}{4}\int_{-\infty}^{\infty} L d\tau = -\rho\frac{\partial\Phi}{\partial z} - \frac{\omega_0}{l\rho} - \frac{c}{2n_o}\rho(\nabla_\perp\Phi)^2 - \frac{\pi^2 c}{24n_o}\frac{(\nabla_\perp\rho)^2}{\rho^3}. \quad (61)$$

Variation of Λ with respect to Φ and ρ leads to the Euler–Lagrange equations

$$\frac{\partial\Phi}{\partial z} + \frac{\mathbf{V}_\perp^2}{2} + \int\frac{dp}{\rho} = \frac{g_0}{\rho^3}\left[\Delta_\perp\rho - \frac{3}{2\rho}(\nabla_\perp\rho)^2\right],$$
$$\frac{\partial\rho}{\partial z} + \nabla_\perp(\rho\mathbf{V}_\perp) = 0, \quad (62)$$

where

$$\mathbf{V}_\perp = \nabla_\perp\Phi, \quad g_0 = \frac{\pi^2 c^2}{12n_o^2},$$

and p is related to ρ by the equation

$$\frac{dp}{d\rho} = \frac{2c\omega_0}{n_o l\rho^2}. \quad (63)$$

In the one-dimensional case ($\nabla_\perp = \Delta_\perp = 0$), system (62), (63) has the solution

$$\rho = \rho_0 = \frac{1}{\tau_p} = \text{const}, \quad \Phi = \frac{z}{v},$$

and v is determined by Eq. (37). Thus, the averaged-Lagrangian method leads to an exact one-soliton solution to Eq. (35) in this case.

The right-hand side of the first equation in (62) represents diffraction effects in transverse dynamics of a pulse. Neglecting the right-hand side, one obtains the eikonal (geometric-optics) approximation for solitons [41]. In this case, system (62) is equivalent to the equations of inviscid fluid dynamics: the former equation is interpreted as the Cauchy theorem for inviscid flows; the latter, as the continuity equation; p, ρ, and $\mathbf{V}_\perp$ correspond to pressure, density, and fluid velocity, respectively. Extending the analogy, one can interpret (63) as the isentropic process associated with fluid motion. Then, transverse stability of the soliton in the eikonal approximation is equivalent to stability of the inviscid flow described by Eqs. (62) and (63), and the corresponding criterion is

$$dp/d\rho > 0.$$

It follows from Eq. (63) that a sine-Gordon soliton is stable in the geometric-optics approximation if the medium is in equilibrium at the initial moment ($1/l \sim W_1 - W_2 > 0$). Expressions (38) and (39) imply that, if

$$dp/d\rho \sim 1/l > 0,$$

then $\mathbf{V}_\perp$ is a monotone increasing function of the USP amplitude. Therefore, wavefront portions characterized by larger amplitudes (which are closer to the center of the USP cross section) move faster than peripheral portions characterized by smaller amplitudes. This defocusing effect may transform the pulse into an "electromagnetic missile" or a "light bullet" [48].

Next, we analyze the effect of diffraction (right-hand side of (62)) on pulse dynamics. System (62) is only slightly different from the system describing the transverse dynamics of a continuous beam [41]. Extending this analogy, we assume that the USP is axially symmetric. Writing Eq. (62) in a cylindrical coordinate system (z, r), we seek a self-similar approximate solution for ρ [41]:

$$\rho(z, r) = \rho_0 \frac{R_0^2}{R^2(z)} \exp\left(-\frac{r^2}{R^2(z)}\right), \quad (64)$$

where R_0 is a constant interpreted as the input USP radius and $R(z)$ as the current radius.

Following [41], we henceforth focus on the paraxial pulse propagation ($r^2/R^2 \ll 1$). Accordingly, we represent the solution for Φ as a series expansion:

$$\Phi(z, r) = f_1(z) + \frac{1}{2} f_2(z) r^2 + \ldots. \quad (65)$$

Substituting (64) and (65) into (62), retaining terms of order up to r^2/R^2, and comparing the coefficients of equal powers of r, we obtain

$$f_2 = R'/R, \quad (66)$$

$$f_2' + f_2^2 = 4QR^2 - 5D, \quad (67)$$

$$f_1' = QR^4 - DR^2, \quad (68)$$

where

$$Q = \frac{\omega_0}{lc/n_0\rho_0^2 R_0^4}, \quad D = \frac{\pi^2 c^2}{3n_0^2\rho_0^2 R_0^4},$$

and primes denote derivatives in z. The last terms in (67) and (68) correspond to deviations from the geometric-optics approximation.

Substituting (66) into (67) leads to an equation for the USP radius:

$$R'' = -\frac{\partial U}{\partial R}, \quad (69)$$

which is formally equivalent to the equation of motion for a Newtonian particle of unit mass in the field with potential energy

$$U(R) = -QR^4 + \frac{5D}{2} R^2. \quad (70)$$

The first integral of Eq. (69) is

$$\frac{R'^2}{2} + U(R) = \Sigma, \quad (71)$$

where the constant

$$\Sigma = -QR_0^4 + 5DR_0^2/2$$

is determined by the input condition $R'(0) = 0$ (according to (66), $f_2(z_0) = 0$ for an input pulse having a plane wave front).

The solution $R(z)$ can be represented in quadratures by integrating Eq. (71). However, a qualitative characterization of the behavior of the USP radius can be obtained by examining the curve of $U(R)$.

The last term on the right-hand side of (70) represents diffraction effects on transverse USP dynamics. This part of $U(R)$ is analogous to the potential energy of the harmonic oscillator, except that $R \geq 0$. Thus, in contrast to the case of a continuous beam with a well-defined carrier frequency, diffraction enhances the USP self-focusing. This distinction can be explained as follows. The dimensionless parameter $\sigma = \lambda/R$, where λ is a characteristic wavelength, determines the effect of wave properties of a pulse on its dynamics. When

$\sigma \ll 1$, the eikonal (geometric-optics) approximation is valid. When $\sigma \sim 1$, wave properties (diffraction) must be allowed for. The wavelength λ of a monochromatic self-focusing beam remains nearly constant while $R \longrightarrow 0$; i.e., σ increases, and diffraction plays an increasingly important role. Therefore, diffraction can cause a self-focused beam to spread out if its intensity is below a certain threshold value [41]. An ultrashort pulse cannot be characterized by a carrier frequency, and the role of λ is played by its size along the direction of pulse propagation:

$$\lambda \sim v\tau = v/\rho,$$

where v is the pulse velocity, which is almost equal to c/n_o in the quasi-unidirectional approximation adopted here. Using (17), we estimate σ at the center of the USP cross section as

$$\sigma \sim \frac{cR}{n_o \rho_0 R_0^2}.$$

This means that $\sigma \longrightarrow 0$ for a self-focusing pulse; i.e., the reliability of the eikonal approximation increases and the relative role played by diffraction decreases.

According to Eq. (69), defocusing corresponds to

$$\left(\frac{\partial U}{\partial R}\right)_{R = R_0} < 0,$$

which yields

$$R_0 > R_c = \pi\sqrt{\frac{5c}{12 n_o a}}. \qquad (72)$$

In the opposite case, a soliton described by the sine-Gordon equation is self-focusing. Note that the eikonal approximation used in [24] reveals only the defocusing effect. Therefore, the conclusion about stability of the soliton with respect to self-focusing made in [24] applies only to pulses with relatively large cross sections. To estimate R_c for a KDP-type ferroelectric, we set $d \sim 10^{-18}$ CGSE units, $N \sim 10^{21}$ cm^{-3}, $\omega_0 \sim 10^{13}$ s^{-1}, and $n_o \approx 1$. Then, $a \sim 10^{15}$ s^{-1} cm^{-1}, and (72) yields $R_c \sim 0.1$ mm.

Equation (45) cannot be associated with any Lagrangian, which makes it impossible to apply the averaged-Lagrangian method. It is clear that (45) reduces to the sine-Gordon equation under the strengthened version of condition (50),

$$\omega_0 \tau_p \gg \left(\frac{D}{2d}\right)^2.$$

In this case, the averaged-Lagrangian method is applicable, and the results concerning the transverse dynamics of the sine-Gordon soliton described above are valid. If the strengthened version of condition (50) is not satisfied, then we can use qualitative arguments valid in the eikonal approximation, as in the case of the sine-Gordon soliton considered above. Since the ordinary component of the USP described by (46)–(48) is dominant, the arguments are developed here for this component. The amplitude $E_{om} = d\varepsilon/\hbar\tau_p$ of the wave described by (46) is a nonmonotonic function of τ_p. With decreasing τ_p, E_{om} monotonically increases as long as

$$\omega_0 \tau_p > \sqrt{2}(D/2d)^2$$

and decreases otherwise. Using (37), we conclude that this inequality is a stability criterion for the USP described by (46)–(48) in the eikonal approximation. Note that this criterion is more restrictive than condition (50).

Finally, we analyze the effect of diffraction on the dynamics of the rational USP considered in Section 5. Substituting the latter ansatz in (53) into the first equation in (52), we obtain

$$\frac{\partial\theta}{\partial z \partial\tau} + \mu\left(\sin\theta - \frac{1}{2}\sin 2\theta\right) = \frac{c}{2n_o}\Delta_\perp\theta, \qquad (73)$$

where

$$\mu = \frac{c\beta_o(\beta_e + \beta_o q^2)}{(n_e - n_o)(1 + q^2)}.$$

The Lagrangian associated with (73) is

$$L = \frac{1}{2}\frac{\partial\theta}{\partial z}\frac{\partial\theta}{\partial\tau} - \mu\sin^2\frac{\theta}{2} + \frac{\mu}{2}\sin^2\theta - \frac{c}{4n_o}(\nabla_\perp\theta)^2. \qquad (74)$$

We seek a trial solution corresponding to (54) in the following form:

$$\theta = -2\,\mathrm{arccot}[\rho(\tau - \Phi)].$$

Substituting it into (74) and integrating the result with respect to τ, we obtain the averaged Lagrangian

$$\Lambda = \frac{1}{\pi}\int_{-\infty}^{+\infty} L d\tau$$

$$= -\rho\frac{\partial\Phi}{\partial z} - \frac{c}{8n_o}\left(\frac{(\nabla_\perp\rho)^2}{\rho^3} + \rho(\nabla_\perp\Phi)^2\right).$$

Variation of Λ with respect to Φ and ρ leads to equa-

tions similar to (62) and (63) with

$$g_0 = \left(\frac{c}{4n_o}\right)^2, \quad \frac{dp}{d\rho} = 0.$$

The latter equality means that the eikonal approximation is not sufficient to determine stability of a rational USP with respect to transverse perturbation. This is explained by the fact that the second and third terms on the left-hand side of (73) cancel out after averaging over τ.

In this case, the procedure proposed above to take into account diffraction effects leads to an equation similar to (69) for the USP radius, where U is given by (70) with

$$Q = 0, \quad D = \frac{c^2}{4n_o^2\rho_0^2R_0^4}.$$

Thus, we have the harmonic-oscillator equation of motion. Under the initial conditions

$$R(0) = R_0, \quad R'(0) = 0$$

it describes the self-focusing of a USP:

$$R = R_0\cos\left(\frac{\pi z}{2l_f}\right), \tag{75}$$

where the self-focusing length l_f is expressed as

$$l_f = \frac{\pi n_o}{\sqrt{5}c\tau_p}R_0^2, \tag{76}$$

and $\tau_p = 1/\rho_0$ is the input USP duration.

Note that l_f is independent of parameters of the tunneling states, because the transverse dynamics is entirely controlled by diffraction, whereas the eikonal approximation is not sufficient to describe the transverse dynamics of a rational USP. Setting $\tau_p \sim 10^{-14}$ s, $n_o \sim 1$, and $R_0 \sim 0.1$ mm, we obtain $l_f \sim 1$ cm.

Thus, diffraction effects are reduced to self-focusing of a rational ultrashort pulse.

7. CONCLUSIONS

This study demonstrates the substantial effect of anisotropy of tunneling on USP dynamics. The system of nonlinear wave equations for the ordinary and extraordinary components, (28) and (29), describes USP propagation at an arbitrary angle to the anisotropy axis. Expressions (30)–(32) determine the behavior of the longitudinal field component. It is important in this context that the longitudinal component can be generated when a USP spanning the tunneling states propagates parallel or perpendicular to the anisotropy axis. It is well known that there is no such component in monochromatic beams and USPs propagating in these directions unless quantum transitions are spanned by the USP spectrum.

A detailed analysis of the regimes of USP propagation parallel and perpendicular to the anisotropy axis is presented. In the former case, self-induced transparency described by the sine-Gordon equation is predicted. This regime is associated with strong excitation of tunneling states. Diffraction leads to self-focusing of the corresponding solitons if their transverse size is less than the critical radius given by (72). In the opposite case, defocusing is predicted. When the direction of propagation is perpendicular to the anisotropy axis, a regime analogous to self-induced transparency can be observed. It is described by Eq. (45), which is different from the sine-Gordon equation. Propagation of a solitary pulse in an array of tunneling states is also associated with their strong excitation. It is shown that the duration of such a soliton-like pulse is limited from below by condition (50). The existence of this limitation essentially depends on the permanent dipole moment D. A qualitative analysis shows that the effect of transverse perturbation strengthens inequality (50).

When the difference between the ordinary and extraordinary refractive indices due to nonresonant processes in the medium is sufficiently large, high-energy resonant soliton-like pulses can develop, with amplitudes decreasing toward infinity as powers of coordinate and time. When the ordinary component is dominant, strong excitation of tunneling states is predicted. In the opposite limit, the populations of the tunneling sublevels remain confined and the propagation velocity is determined solely by the extraordinary refractive index.

Rational soliton-like pulses can develop in an array of tunneling states if the integrated area A_0 of the incoming signal exceeds 2π. If $A_0 < 2\pi$, then the pulse decays.

Note that the transverse dynamics of resonant USPs considered here is entirely determined by their wave properties (diffraction), whereas the geometric-optics stage corresponds to "neutral equilibrium." Therefore, the self-focusing length l_f is determined by the input USP parameters, being totally independent of tunneling characteristics (see (76)). Note that l_f is proportional to the squared input USP radius and inversely proportional to its duration. When $\tau_p \sim 10$ fs and $R_0 \sim 0.1$ mm, we obtain $l_f \sim 1$ cm. By increasing the radius to 1 mm, the self-focusing length can be increased by a factor of 100, i.e., to 1 m. This makes it possible to conduct experiments on rational soliton-like pulses. Reduction of τ_p to several femtoseconds would broaden the USP spectrum and, therefore, make it necessary to allow for optical electronic transitions and take into account higher lying proton states in double-well crystal poten-

tials. In the first approximation, the ensuing effects can be treated as weak because of weak interaction of the USP spectrum with the corresponding states and can be taken into account in the optical-transparency approximation [49], while optical electronic transitions can be taken into account additively.

As an example of tunneling, we mentioned proton transitions in KDP-type ferroelectrics. Note that the current progress in nanotechnologies makes it possible to deal with coherent optical effects in arrays of quantum dots and/or quantum wells in semiconductors [50–52], where the tunneling particles are electrons rather than protons. Since semiconductor structures are low dimensional, these tunneling states are sufficiently anisotropic for permanent dipole moments to exist [28, 30]. Therefore, we can hope to reveal USP propagation regimes similar to those described in this study in strongly anisotropic, artificially grown semiconductor crystals.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 02-02-17710a.

REFERENCES

1. É. M. Belenov, P. G. Kryukov, A. V. Nazarkin, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **47**, 442 (1988) [JETP Lett. **47**, 523 (1988)].
2. É. M. Belenov and A. V. Nazarkin, Pis'ma Zh. Éksp. Teor. Fiz. **51**, 252 (1990) [JETP Lett. **51**, 288 (1990)].
3. É. M. Belenov, A. V. Nazarkin, and V. A. Ushchapovskiĭ, Zh. Éksp. Teor. Fiz. **100**, 762 (1991) [Sov. Phys. JETP **73**, 422 (1991)].
4. A. I. Maĭmistov and S. O. Elyutin, Opt. Spektrosk. **70**, 101 (1991) [Opt. Spectrosc. **70**, 57 (1991)].
5. A. I. Maĭmistov, Opt. Spektrosk. **76**, 636 (1994) [Opt. Spectrosc. **76**, 569 (1994)].
6. A. I. Maĭmistov, Opt. Spektrosk. **78**, 483 (1995) [Opt. Spectrosc. **78**, 435 (1995)].
7. S. V. Sazonov and E. V. Trifonov, J. Phys. B: At. Mol. Opt. Phys. **27**, L7 (1994).
8. S. V. Sazonov, Zh. Éksp. Teor. Fiz. **107**, 20 (1995) [JETP **80**, 10 (1995)].
9. S. A. Kozlov and S. V. Sazonov, Zh. Éksp. Teor. Fiz. **111**, 404 (1997) [JETP **84**, 221 (1997)].
10. A. I. Maĭmistov, Kvantovaya Élektron. (Moscow) **30**, 287 (2000).
11. S. A. Kozlov, in *Problems in Coherent and Nonlinear Optics* (ITMO, St. Petersburg, 2000), p. 12 [in Russian].
12. T. Brabec and F. Krausz, Rev. Mod. Phys. **72**, 545 (2000).
13. S. V. Sazonov, Zh. Éksp. Teor. Fiz. **119**, 419 (2001) [JETP **92**, 361 (2001)].
14. S. V. Sazonov, Usp. Fiz. Nauk **171**, 663 (2001) [Phys. Usp. **44**, 631 (2001)].
15. S. V. Sazonov and A. F. Sobolevskiĭ, Pis'ma Zh. Éksp. Teor. Fiz. **75**, 746 (2002) [JETP Lett. **75**, 621 (2002)].
16. V. G. Bespalov, S. A. Kozlov, A. Yu. Shpolyansky, and I. Walmsley, Phys. Rev. A **66**, 013811 (2002).
17. A. M. Zheltikov, Usp. Fiz. Nauk **172**, 743 (2002) [Phys. Usp. **45**, 687 (2002)].
18. S. V. Sazonov and A. F. Sobolevskiĭ, Zh. Éksp. Teor. Fiz. **123**, 919 (2003) [JETP **96**, 807 (2003)].
19. S. A. Akhmanov, V. A. Vysloukh, and A. S. Chirkin, *The Optics of Femtosecond Laser Pulses* (Nauka, Moscow, 1988) [in Russian].
20. S. L. McCall and L. Hahn, Phys. Rev. Lett. **18**, 908 (1967).
21. G. L. Lamb, Rev. Mod. Phys. **43**, 99 (1971).
22. A. I. Maimistov and A. M. Basharov, *Nonlinear Optical Waves* (Kluwer Academic, Dordrecht, 1999).
23. A. Yu. Parkhomenko and S. V. Sazonov, Zh. Éksp. Teor. Fiz. **114**, 1595 (1998) [JETP **87**, 864 (1998)].
24. S. V. Nesterov and S. V. Sazonov, Fiz. Tverd. Tela (St. Petersburg) **45**, 303 (2003) [Phys. Solid State **45**, 319 (2003)].
25. B. A. Strukov and A. P. Levanyuk, *Physical Principles of Ferroelectric Phenomena in Crystals* (Nauka, Moscow, 1995) [in Russian].
26. V. G. Vaks, *Introduction to the Microscopic Theory of Ferroelectrics* (Nauka, Moscow, 1973) [in Russian].
27. S. V. Sazonov, Fiz. Tverd. Tela (St. Petersburg) **37**, 1626 (1995) [Phys. Solid State **37**, 875 (1995)].
28. S. Kocinac, Z. Ikonic, and V. Milanovic, Opt. Commun. **140**, 89 (1997).
29. M. Agrotis, N. M. Ercolani, S. V. Glasgow, and J. V. Moloney, Physica D (Amsterdam) **138**, 134 (2000).
30. A. A. Zabolotskiĭ, Zh. Éksp. Teor. Fiz. **121**, 1012 (2002) [JETP **94**, 869 (2002)].
31. A. I. Maĭmistov and J.-G. Caputo, Opt. Spektrosk. **94**, 275 (2003) [Opt. Spectrosc. **94**, 245 (2003)].
32. S. V. Sazonov, Zh. Éksp. Teor. Fiz. **124**, 803 (2003) [JETP **97**, 722 (2003)].
33. M. Born and E. Wolf, *Principles of Optics*, 4th ed. (Pergamon Press, Oxford, 1969; Nauka, Moscow, 1973).
34. A. Yu. Parkhomenko and S. V. Sazonov, Kvantovaya Élektron. (Moscow) **27**, 139 (1999).
35. S. V. Voronkov and S. V. Sazonov, Zh. Éksp. Teor. Fiz. **120**, 269 (2001) [JETP **93**, 236 (2001)].
36. S. V. Sazonov, Opt. Spektrosk. **95**, 656 (2003) [Opt. Spectrosc. **95**, 622 (2003)].
37. N. N. Moiseev, *Asymptotic Methods of Nonlinear Mechanics* (Nauka, Moscow, 1981) [in Russian].
38. A. H. Nayfeh, *Introduction to Perturbation Techniques* (Wiley, New York, 1981; Mir, Moscow, 1984).
39. V. A. Yakubovich and V. M. Starzhinskiĭ, *Parametric Resonance in Linear Systems* (Nauka, Moscow, 1987) [in Russian].
40. J. D. Gibbon, P. J. Gaudrey, J. K. Elibek, and R. K. Bullough, J. Phys. A: Math. Gen. **6**, 1237 (1973).
41. N. V. Karlov and N. A. Kirichenko, *Oscillations, Waves, and Structures* (Nauka, Moscow, 2001) [in Russian].

42. A. I. Maĭmistov, Kvantovaya Élektron. (Moscow) **11**, 567 (1984).

43. L. A. Bol'shov and V. V. Likhanskiĭ, Kvantovaya Élektron. (Moscow) **12**, 1339 (1985).

44. A. A. Zabolotskiĭ, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 709 (2002) [JETP Lett. **76**, 607 (2002)].

45. A. A. Zabolotskiĭ, Zh. Éksp. Teor. Fiz. **123**, 560 (2003) [JETP **96**, 496 (2003)].

46. S. V. Voronkov and S. V. Sazonov, Fiz. Tverd. Tela (St. Petersburg) **43**, 1969 (2001) [Phys. Solid State **43**, 2051 (2001)].

47. S. K. Zhdanov and B. A. Trubnikov, *Quasi-Gaseous Unstable Media* (Nauka, Moscow, 1991) [in Russian].

48. D. E. Edmundson and R. H. Enns, Phys. Rev. A **51**, 2491 (1995).

49. S. V. Sazonov and A. F. Sobolevskiĭ, Opt. Spektrosk. **90**, 449 (2001) [Opt. Spectrosc. **90**, 390 (2001)].

50. V. Ya. Demikhovskiĭ and G. A. Vugal'ter, *Physics of Quantum Low-Dimensional Structures* (Logos, Moscow, 2000) [in Russian].

51. A. I. Maĭmistov, Opt. Spektrosk. **93**, 49 (2002) [Opt. Spectrosc. **93**, 45 (2002)].

52. A. I. Maĭmistov and S. O. Elyutin, Opt. Spektrosk. **93**, 274 (2002) [Opt. Spectrosc. **93**, 257 (2002)].

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 3, 2004, pp. 663–668.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 3, 2004, pp. 758–764.

NONLINEAR PHYSICS

Chaos in a Driven Pendulum under Asymmetric Forcing

V. V. Vecheslavov

Budker Institute of Nuclear Physics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia
e-mail: vecheslavov@inp.nsk.su

Received November 14, 2003

Abstract—An analysis of the stochastic layer in a pendulum driven by an asymmetric high-frequency perturbation of fairly general form is continued. Analytical expressions are found for the amplitudes of secondary harmonics, and their contributions to the amplitude of the separatrix map responsible for onset of dynamical chaos are evaluated. Additional evidence is presented of the previously established fact that the secondary harmonics completely determine the stochastic-layer width when the primary frequencies lie in certain intervals. The mechanism of the onset of chaos in the vicinity of zeros of Melnikov integrals is shown to be substantially different as compared to the previously analyzed case of symmetric perturbation. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The phase space of a typical (i.e., nonintegrable) Hamiltonian system consists of chaotic and regular regions. It is well known that interaction between nonlinear resonances is the mechanism responsible for onset of dynamical chaos. Generally, initial conditions are set in the neighborhood of one resonance (the "main" one), and the remaining resonances are treated as a perturbation. The most interesting—and often unexpected—dynamics are observed in the vicinity of the separatrices of the main resonance, which separate regions where the phase rotates (outside resonance) from those where it oscillates (inside resonance).

The neighborhoods of separatrices have been commonly considered to be the "origin" of chaos, because motion along separatrices is characterized by infinitely large periods and substantial interaction between resonances. However, this is true only when the main resonance corresponds to an analytic potential with exponentially decreasing Fourier amplitudes. In this case, each separatrix is split by perturbation into two distinct trajectories. Trajectories of this kind fill a narrow region and make up a stochastic layer [1–4], in which three parts should be distinguished: an upper one of width w_u (where the phase x rotates so that $p > 0$), a middle one of width w_m (where the phase oscillates), and a lower one of width w_l (where the phase x rotates so that $p < 0$).

Qualitatively different behavior is observed when the potential is a smooth function with Fourier amplitudes decreasing as a power of the harmonic number. Some striking examples of preserved integer- and fractional-resonance separatrices and absence of stochastic layer in their neighborhoods were discussed in [5–7] for perturbed piecewise linear systems. It should be emphasized that these systems remain nonintegrable and their separatrices are preserved despite the presence of strong local chaos (see Fig. 4 in [6]).

Furthermore, it was found that the onset of chaos strongly depends on the spectrum of the perturbation. Let us briefly recall the history of the problem.

The first analysis of a system subject to a symmetric high-frequency perturbation was presented by Chirikov [1]. He showed that both amplitude of the separatrix map and energy width of the stochastic layer exponentially decrease with increasing frequency and the three parts of the layer have equal widths in the high-frequency limit.

In a recent study [8], the low-frequency asymptotic behavior of the same system was considered. In this limit, it was found that the separatrix-map amplitude linearly increases with frequency, while the layer width is independent of frequency. Both asymptotics are relatively simple, and the problem is most difficult to analyze in the intermediate frequency range, where the adiabaticity parameter cannot be treated as small or large. It was noted that the so-called resonance invariants could play an important role in this range, since they adequately represent the topology of individual resonances. Resonance invariants of the first to third order (corresponding to the 1 : 1, 1 : 2, and 1 : 3 resonances) were found for Chirikov's standard map in [9] and for the single-frequency separatrix map in [10]. A sufficiently detailed characterization of the structure of the stochastic layer was recently presented in [11] for a pendulum driven by symmetric perturbation of arbitrary frequency.

The first analysis of asymmetric perturbation was presented in [12, 13], where the Hamiltonian of a pendulum driven by two harmonics:

$$H(x, p, t) = \frac{p^2}{2} + \cos x + V(x, t), \tag{1}$$

$$V(x, t) = \varepsilon_1 \cos(x - \Omega_1 t) + \varepsilon_2 \cos(x - \Omega_2 t). \tag{2}$$

The harmonic amplitudes were assumed to be small (ε_1, $\varepsilon_2 \ll 1$), while the frequencies were assumed to be high relative to the natural oscillation frequency ($|\Omega_1|$, $|\Omega_2| \gg 1$).

It was found that, in addition to the primary frequencies Ω_1 and Ω_2, the perturbed motion has secondary harmonics with amplitudes proportional to $\varepsilon_1\varepsilon_2$ and frequencies equal to the sum and difference of the primary frequencies. Moreover, it was found that the secondary harmonics are much weaker than the primary ones when $\varepsilon_1, \varepsilon_2 \ll 1$.

Even the first numerical experiments revealed the seemingly surprising fact that these secondary harmonics completely determine both separatrix-map amplitude and stochastic-layer width under certain conditions [12, 13]. In the example of system (1), (2) considered in [13], the parameter values were $\varepsilon_1 = \varepsilon_2 = 0.075$, $\Omega_1 = 13$, and $\Omega_2 = -10$. The amplitude of the sum-frequency harmonic of the perturbed motion (with $\Delta\Omega_+ = 3$) was $\varepsilon \approx 4.5 \times 10^{-5}$ [12], which is smaller than the primary-harmonic amplitude by a factor of about 1700. However, its contribution to the separatrix-map amplitude corresponding to the upper part of the layer (responsible for the onset of chaos) exceeded the total contribution of the primary harmonics by a factor of almost 400, while the individual parts of the layer had substantially different widths. This is explained by the exponential dependence of layer width on frequency at $\Omega \gg 1$, which leads to a dominant effect of weak low-frequency harmonics on the onset of chaos. The important role played by sum-frequency secondary harmonics was also demonstrated for smooth systems [14].

The sum-frequency secondary harmonic of the perturbed motion has the form

$$\varepsilon_+\cos(2x - \Delta\Omega_+ t), \quad \varepsilon_+ = -\frac{\varepsilon_1\varepsilon_2}{2}\left[\frac{1}{\Omega_1^2} + \frac{1}{\Omega_2^2}\right], \quad \Delta\Omega_+ = \Omega_1 + \Omega_2 > 0 \tag{3}$$

and gives rise to a harmonic with the same frequency and the amplitude

$$W_+ = \frac{4\pi}{3}\varepsilon_1\varepsilon_2\left[\frac{1}{\Omega_1^2} + \frac{1}{\Omega_2^2}\right] \times \frac{\exp(\pi\Delta\Omega_+/2)}{\sinh(\pi\Delta\Omega_+)}\Delta\Omega_+^2(\Delta\Omega_+^2 - 2). \tag{4}$$

The perturbed motion has two symmetric difference-frequency harmonics:

$$\varepsilon_-[\cos(x - \Delta\Omega_- t) - \cos(x + \Delta\Omega_- t)],$$

$$\varepsilon_- = \frac{\varepsilon_1\varepsilon_2}{4}\left[\frac{1}{\Omega_1^2} - \frac{1}{\Omega_2^2}\right]\frac{1}{\Delta\Omega_-^2}, \tag{5}$$

$$\Delta\Omega_- = \Omega_2 - \Omega_1 > 0,$$

which give rise to a separatrix-map harmonic with the frequency $\Delta\Omega_-$ and the amplitude

$$W_- = -\frac{\pi\varepsilon_1\varepsilon_2}{\cosh(\pi\Delta\Omega_-/2)}\left[\frac{1}{\Omega_1^2} - \frac{1}{\Omega_2^2}\right]. \tag{6}$$

In this paper, the analysis of the onset of chaos in a pendulum driven by a high-frequency perturbation that has a more complicated form as compared to (2):

$$V(x, t) = \varepsilon_1\cos(m_1 x - \Omega_1 t) + \varepsilon_2\cos(m_2 x - \Omega_2 t), \tag{7}$$

where m_1 and m_2 are arbitrary positive integer numbers. Note that these parameters determine the structure (in particular, the number of zeros) of Melnikov integrals. The characteristics of the onset of dynamical chaos in the vicinity of these zeros are discussed in Section 3.

2. AMPLITUDES OF SECONDARY HARMONICS

Consider the Hamiltonian of a pendulum driven by the single-frequency perturbation

$$V(x, t) = \varepsilon\cos\left(\frac{n}{2}x - \Omega t\right) \tag{8}$$

with positive n and Ω. Since this perturbation harmonic is a resonance lying above the main resonance in the phase plane, it is called the upper harmonic for convenience.

It was shown by Chirikov in [1] that the amplitude of the separatrix map for the upper part of the stochastic layer generated by upper harmonic (8) is expressed as

$$W_T(\Omega, n) = \varepsilon\Omega A_n(\Omega) \tag{9}$$

in terms of the Melnikov integrals

$$A_n(\Omega > 0) = \frac{2\pi}{(n-1)!}\frac{\exp(\pi\Omega/2)}{\sinh(\pi\Omega)} \times (2\Omega)^{n-1}[1 + f_n(\Omega)], \tag{10}$$

$$f_1 = f_2 = 0, \quad f_{n+1} = f_n - (1 + f_{n-1})\frac{n(n-1)}{4\Omega^2}, \quad n \geq 3. \tag{11}$$

The parameter n is called the index of Melnikov integral here.

By replacing Ω with $-\Omega$ in (8), the upper harmonic is transformed into the lower one. The corresponding contribution to the amplitude of the separatrix map for the upper part of the layer is evaluated by using an essentially different expression in (9):

$$A_n(\Omega < 0) = (-1)^n A_n(|\Omega|)\exp(-\pi|\Omega|). \tag{12}$$

It should be noted that the derivation of (9)–(12) in [1] did not rely on any simplifying assumptions, and these relations are valid for $0 < |\Omega| < \infty$. Note also that integrals (10) considered here have even indices n, because perturbation (7) contains only integer m_1 and m_2.

Let us estimate the secondary-harmonic amplitudes, which are not known a priori. Following [12], change from the coordinate $x(t)$ and momentum $p(t)$ to the deviations from their values $x_s(t) = 4\arctan e^t$ and $p_s(t) = 2\sin x_s(t)$ on the unperturbed separatrix,

$$y(t) = x(t) - x_s(t), \quad u(t) = p(t) - p_s(t), \qquad (13)$$

and use the generating function

$$F_2(u, x, t) = [p_s(t) - u][x - x_s(t)]$$

to rewrite the Hamiltonian (1), (7) as follows:

$$\begin{aligned} H(y, u, t) &= \frac{u^2}{2} + \cos y \cos x_s(t) \\ &\quad - \sin y \sin x_s(t) + y \sin x_s(t) \\ &\quad + \sum_{k=1}^{2} \varepsilon_k [\cos(m_k y)\cos(m_k x_s(t) - \Omega_k t) \\ &\quad - \sin(m_k y)\sin(m_k x_s(t) - \Omega_k t)]. \end{aligned} \qquad (14)$$

Since the perturbation is weak, assume that $|y(t)| \ll 1$ and replace $\cos(my) \longrightarrow 1 - (my)^2/2$ and $\sin(my)$ with my to derive the equation

$$\begin{aligned} \frac{d^2y}{dt^2} &= y\left[\cos x_s + \sum_{k=1}^{2} \varepsilon_k m_k^2 \cos(m_k x_s - \Omega_k t)\right] \\ &\quad + \sum_{k=1}^{2} \varepsilon_k m_k \sin(m_k x_s - \Omega_k t). \end{aligned}$$

Define Δy_ε as the difference between the left- and right-hand sides of this equation:

$$\begin{aligned} \Delta y_\varepsilon &= \frac{d^2y}{dt^2} - y\left[\cos x_s + \sum_{k=1}^{2} \varepsilon_k m_k^2 \cos(m_k x_s - \Omega_k t)\right] \\ &\quad - \sum_{k=1}^{2} \varepsilon_k m_k \sin(m_k x_s - \Omega_k t). \end{aligned} \qquad (15)$$

The forced solution y_ε of interest here (vanishing for $\varepsilon = 0$) can be found by performing successive approximations in order to reduce Δy_ε to zero [12]. After two iterative steps, the result is

$$\begin{aligned} y_\varepsilon^{(2)}(t) &\approx -\sum_{k=1}^{2} \frac{\varepsilon_k m_k}{(m_k p_s - \Omega_k)^2} \sin(m_k x_s - \Omega_k t) \\ &\quad - \frac{\varepsilon_1 \varepsilon_2 m_1 m_2}{2}\left\{\left[\frac{m_2}{(m_1 p_s - \Omega_1)^2} + \frac{m_1}{(m_2 p_s - \Omega_2)^2}\right]\right. \\ &\quad \times \frac{\sin(m_+ x_s - \Delta\Omega_+ t)}{(m_+ p_s - \Delta\Omega_+)^2} \\ &\quad + \left[\frac{m_1}{(m_2 p_s - \Omega_2)^2} - \frac{m_2}{(m_1 p_s - \Omega_1)^2}\right] \\ &\quad \left.\times \frac{\sin(m_- x_s - \Delta\Omega_- t)}{(m_- p_s - \Delta\Omega_-)^2}\right\} + \ldots, \end{aligned} \qquad (16)$$

where unimportant terms are omitted and $m_+ = m_1 + m_2$ and $m_- = m_2 - m_1$ are introduced by analogy with the sum and difference frequencies.

By virtue of the inequality $|\Omega_{1,2}| \gg p_{s,\max} \approx 2$, the terms mp_s in the denominators in (16) that contain these frequencies can be dropped to simplify further expressions. The denominators that contain the sum and difference frequencies are simplified similarly without any substantiation. However, the ensuing errors are corrected below by introducing empirical numerical coefficients into the final results (see Section 3).

Returning to (1) and (7), set $x = x_s(t) + y_\varepsilon^{(2)}(t)$. Since motion in the neighborhood of the unperturbed separatrix is considered, the replacements $\cos(my) \longrightarrow 1$ and $\sin(my) \approx my_\varepsilon^{(2)}$ can be used to rewrite perturbation (7) as

$$\begin{aligned} &V(y, t) \\ &\approx -y_\varepsilon^{(2)}\left[\sin x_s + \sum_{k=1}^{2} \varepsilon_k m_k \sin(m_k x_s - \Omega_k t)\right]. \end{aligned} \qquad (17)$$

Substituting (16) into (17), one finds that sum- and difference-frequency harmonics (which are the only important ones in this analysis) can arise in the perturbed system in two ways. First, the sum in (16) interfere with the primary harmonics in (17), giving rise to a sum-frequency harmonic,

$$\begin{aligned} &\varepsilon_+ \cos(m_+ x_s - \Delta\Omega_+ t), \\ &\varepsilon_+ = -\frac{\varepsilon_1 \varepsilon_2 m_1 m_2}{2}\left[\frac{m_2}{\Omega_1^2} + \frac{m_1}{\Omega_2^2}\right], \end{aligned} \qquad (18)$$

and a difference-frequency one,

$$\varepsilon_- \cos(m_- x_s - \Delta\Omega_- t),$$
$$\varepsilon_- = \frac{\varepsilon_1\varepsilon_2 m_1 m_2}{2}\left[\frac{m_2}{\Omega_1^2} - \frac{m_1}{\Omega_2^2}\right]. \quad (19)$$

Second, the terms proportional to $\varepsilon_1\varepsilon_2$ in (16) interfere with $\sin x_s$ in (17), giving rise to two sum-frequency harmonics,

$$\frac{\varepsilon_+}{2\Delta\Omega_+^2}[\cos((m_+ - 1)x_s - \Delta\Omega_+ t) - \cos((m_+ + 1)x_s - \Delta\Omega_+ t)], \quad (20)$$

where ε_+ is defined in (18), and two difference-frequency ones,

$$\frac{\varepsilon_-}{2\Delta\Omega_-^2}[\cos((m_- - 1)x_s - \Delta\Omega_- t) - \cos((m_- + 1)x_s - \Delta\Omega_- t)], \quad (21)$$

where ε_- is defined in (19).

Once the amplitudes of secondary harmonics are known, relation (9) can be used to express their normalized amplitudes, $W^* = W/\varepsilon_1\varepsilon_2$, in terms of Melnikov integrals (10):

$$W_\pm^* = \frac{W_\pm}{\varepsilon_1\varepsilon_2} = \tilde{\varepsilon}_\pm\left[\Delta\Omega_\pm A_{2m_\pm}(\Delta\Omega_\pm) + \frac{A_{2m_\pm - 2}(\Delta\Omega_\pm) - A_{2m_\pm + 2}(\Delta\Omega_\pm)}{2\Delta\Omega_\pm}\right], \quad (22)$$

$$\tilde{\varepsilon}_\pm = \frac{a_\pm m_1 m_2}{2}\left[\frac{m_2}{\Omega_1^2} \pm \frac{m_1}{\Omega_2^2}\right]. \quad (23)$$

Here, the upper and lower signs in the subscripts correspond to the sum and difference frequencies, respectively. In (23), empirical coefficients a_+ and a_- are introduced. Practical application of (22) has shown that the first term in brackets plays a dominant role.

In the next section, the results of numerical verification of (22) and (23) are presented and some interesting details of the onset of chaos are discussed.

3. COMPARISON OF THEORETICAL RESULTS WITH NUMERICAL EXPERIMENT

The separatrix map originally introduced in [15] characterizes the behavior of a dynamical system in the neighborhood of a separatrix. For pendulum, it is defined as

$$\bar{w} = w + \sum_l W_l \sin(\Omega_l t_\pi),$$
$$\bar{t}_\pi = t_\pi + \ln\frac{32}{|\bar{w}|}, \quad l = 1, 2, \ldots, \quad (24)$$

where $w = H(x, p, t) - 1$ is the relative energy deviation from the unperturbed separatrix, t_π denotes the moments when the system passes through a stable equilibrium point $x = \pi$. The sum must contain all harmonics that are important for the analyzed part of the stochastic layer, both primary ones (contained in (7)) and secondary ones (not contained in (7)).

When the frequencies Ω_l are incommensurate, the moments t_π are measured on a continuous time scale. When the frequencies are multiples of some reference frequency Ω_0, the last relation in (24) may be rewritten as

$$\bar{\psi}_\pi = \psi_\pi + \Omega_0 \ln\frac{32}{|\bar{w}|}, \quad \psi_\pi = \Omega_0 t_\pi \bmod 2\pi. \quad (25)$$

Iterating the map is the fastest method for evaluating the widths of individual parts of the stochastic layer, and this justifies the effort required to construct the map.

Before comparing theoretically and numerically evaluated separatrix-map amplitudes, the numerical algorithm for constructing the map is briefly recalled (see [12] for details).

First of all, the central homoclinic point p_{fb} is found with high accuracy on the symmetry line $x = \pi$, as a boundary between the regions of rotating and oscillating phase. Then, a trajectory is constructed starting from an arbitrary point in a narrow momentum interval on the line $x = \pi$, $p_{fb} < p < p_{fb} + \delta p$, which definitely lies in the analyzed part of the layer. The trajectory is either computed until it spans a prescribed number of periodic orbits or terminated when it enters a different part of the layer. In both cases, a new arbitrary trajectory is constructed starting from the same interval until a required number N of periodic orbits have been traversed. For each period, the energy deviation from the separatrix is evaluated:

$$w = 32\exp(-T), \quad (26)$$

where T is the time interval between two consecutive moments when the system passes through the stable equilibrium point $x = \pi$. Calculating the energy change $\delta w = \bar{w} - w$ between each pair of consecutive periods and assigning it to the moment t_π separating the periods, one can construct separatrix map (24) in terms of $(\delta w)_k$ and $t_{\pi,k}$ ($k = 1, 2, \ldots, N-1$). To be specific, we analyze the upper part of the layer here. Note that the outer (upper and lower) parts of the layer are of particular interest, because they contribute to the overlap of adjacent resonances and the onset of global chaos.

In this study, we consider perturbation (7) with the constant parameters

$$\varepsilon_1 = \varepsilon_2 = 0.05, \quad m_1 = 1, \quad m_2 = 3 \quad (27)$$

for different values of the primary frequencies Ω_1 and Ω_2. Note that the contributions of the primary harmon-

ics to the separatrix-map amplitude is determined, according to (9), by the Melnikov integrals with $n_1 = 2m_1 = 2$ and $n_2 = 2m_2 = 6$, respectively, while $n_- = 2(m_2 - m_1) = 4$ and $n_+ = 2(m_1 + m_2) = 8$ for the difference- and sum-frequency secondary harmonics, respectively.

Apparently, the most important result of the study of perturbation (7) of general form is that it provides additional evidence of the fact (established in [12, 13]) that there exist relatively large intervals of parameters where secondary harmonics play a dominant role in the onset of chaos. Consider system (1), (7), (27) with $\Omega_1 = 15$ and $\Omega_2 = -12$. These frequencies are commensurate, and $\Omega_0 = 3.0$ can be used as a reference frequency in (25). The corresponding amplitudes of the separatrix map constructed for the upper part of the layer by the algorithm outlined above are

$$W(3) \approx 1.35 \times 10^{-4}, \quad W(15) \approx 2.09 \times 10^{-7}, \quad W(-12) \approx 6.18 \times 10^{-7}. \tag{28}$$

In the example considered here, the contribution of the secondary harmonic with the sum frequency $\Delta\Omega_+ = 3$ is greater than those due to the primary harmonics by order of magnitude, and the map for the upper part of the layer can be considered, to high accuracy, as a single-frequency one (with low frequency $\Omega = 3$). It may seem that the only role played by the weak forcing harmonics in the onset of chaos is to give rise to a strong secondary harmonic in the separatrix map.

According to (10) and (11), the Melnikov integrals with $n = 3$ and higher, including the factor in brackets, change their signs as they pass through zero. The integral A_2 has no zeros at all; A_4 has a single zero at $\Omega_1^{(4)} = \sqrt{2}$; A_6 has zeros at $\Omega_1^{(6)} \approx 1.1514\ldots$ and $\Omega_2^{(6)} \approx 2.9452\ldots$; A_8, at $\Omega_1^{(8)} \approx 1.0248\ldots$, $\Omega_2^{(8)} \approx 2.4495\ldots$, and $\Omega_3^{(8)} \approx 4.5771\ldots$; and so on. By analogy with the forcing harmonics, it is convenient to make a distinction between primary and secondary integrals in (10). Note that the four integrals in the example considered here (two primary and two secondary ones) differ by the number of zeros.

Figure 1 illustrates the behavior of the primary and secondary integrals, $A_2(\Omega)$ and $A_8(\Omega)$, and the corresponding normalized separatrix-map amplitudes $W_T^* = W/\varepsilon$. In particular, the linear growth at $\Omega \ll 1$ and exponential decay at $\Omega \gg 1$ of $W_T^*(\Omega)$ calculated theoretically above are demonstrated.

If the secondary integral $A_8(\Omega_+)$ reflects any reality, then the frequency dependence of the separatrix-map amplitude, $W^*(\Delta\Omega_+)$, must have the three zeros characteristic only of this integral. Indeed, consider the case of $\Omega_1 = 14.0$ and $\Omega_2 = $ var illustrated by Fig. 2, where the

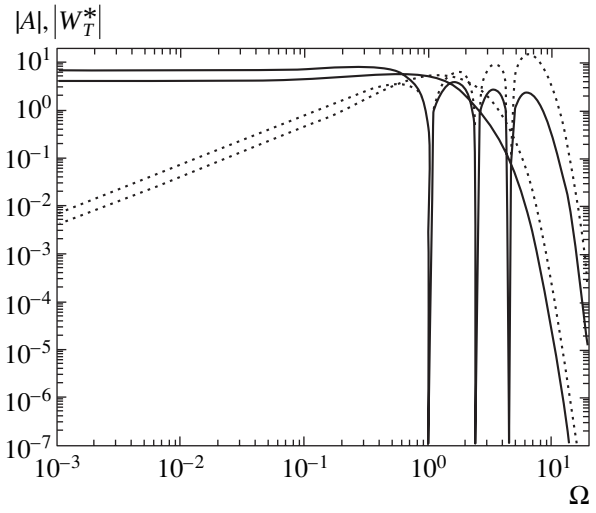

Fig. 1. Absolute values of the Melnikov integrals A_2 and A_8 (solid curves) and the corresponding normalized amplitudes $W_T^* = W_T/\varepsilon$ given by (9) (dotted curves).

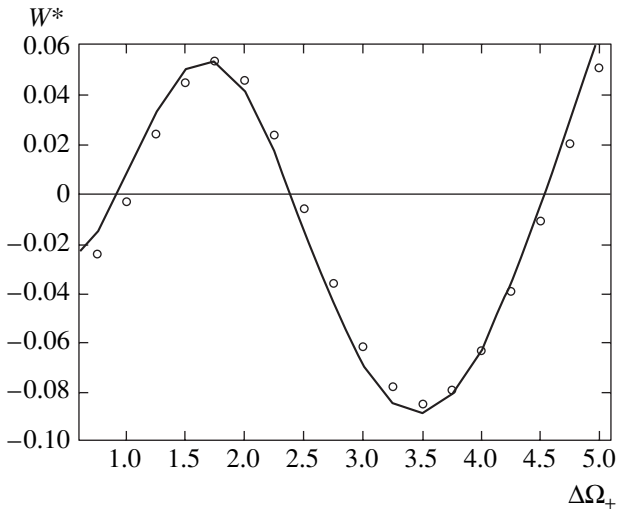

Fig. 2. Normalized amplitude of the sum-frequency secondary harmonic in the separatrix map for $\Omega_1 = 14.0$ and $\Omega_2 =$ var. The solid curve is plotted by using (22) and (23) with empirical coefficient $a_+ \approx 0.25$. Circles represent numerical results obtained by computing the map defined by (24).

theoretical results obtained by using expressions (22) and (23) are represented by the solid curve and the experimental results obtained by computing separatrix map (24) are plotted as circles. It is clear that both functions have zeros almost where expected. However, the meaning of the corresponding empirical coefficient, $a_+ \approx 0.25$, is unclear.

A substantial difference in the behavior of secondary harmonics in the separatrix map considered here from the case of symmetric perturbation analyzed in [8] is localized in the vicinity of zeros of Melnikov integrals. It was found in [8] that the separatrix map has two

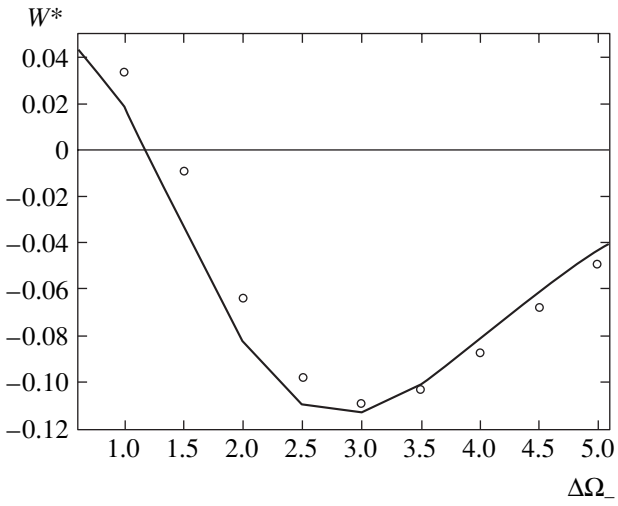

Fig. 3. Results analogous to those in Fig. 2 obtained for the difference-frequency secondary harmonic in the case of $\Omega_1 = 13.0$ and $\Omega_2 = \text{var}$ with $a_- \approx 0.57$.

harmonics in the intermediate frequency range: a single-frequency one and a double-frequency one (the latter corresponds to the difference-frequency secondary harmonic). At the zeros of the integrals in (10), the single-frequency harmonic amplitude passes through zero (its sign changes), but the amplitude of the map remains finite because of the double-frequency contribution. Resonance invariants for the double-frequency harmonic were found in [11] in order to examine the vicinity of zeros of Melnikov integrals.

When the perturbation is asymmetric, no double-frequency harmonic is generated, the amplitude of the separatrix map for the upper part of the stochastic layer is small at the zero of (10), and so is the width of this part of the layer. For example, the energy widths of the individual parts of the layer obtained by using (26) for the shortest period when $\Omega_1 = 14.0$ and $\Omega_2 = -12.9752$ ($\Delta\Omega_+ = 1.0248$ is very close to the first zero of the secondary integral $A_8(\Omega)$) are as follows:

$$w_u \approx 9.72 \times 10^{-7}, \quad w_m \approx 2.51 \times 10^{-2}, \quad w_l \approx 1.66 \times 10^{-2}, \tag{29}$$

where w_u, w_m, and w_l are the widths of the upper, middle, and lower parts of the layer. The upper part (to which the secondary harmonic is "tuned") is much smaller than the remaining ones. Note that a secondary harmonic was intentionally used in [13] to reduce the width of a certain part of the stochastic layer.

The difference-frequency secondary harmonic in the separatrix map is represented by the integral $A_4(\Delta\Omega_-)$, which has the single zero $\Omega_1^{(4)} = \sqrt{2}$. Figure 3 compares theoretical results obtained by using expressions (22) and (23) (solid curve) with numerical experiment (circles). Here, quantitative agreement is not as good as in the preceding example, but the trends exhibited by the theoretical and numerical results are qualitatively similar. The corresponding empirical coefficient is $a_- \approx 0.57$.

4. CONCLUSIONS

The results presented above (see also [12, 13]) make it clear that the secondary harmonics corresponding to the sum and difference of the primary frequencies (contained in the Hamiltonian of the system) are real entities that contribute decisively to the formation of the stochastic layer under certain conditions. Therefore, the development of a comprehensive theory of dynamical chaos in Hamiltonian systems must include their further detailed experimental and theoretical analysis.

ACKNOWLEDGMENTS

This work was supported, in part, by the Russian Foundation for Basic Research, project no. 01-02-16836, and by the Russian Academy of Sciences under the multidisciplinary scientific program "Mathematical Methods in Nonlinear Mechanics."

REFERENCES

1. B. V. Chirikov, Phys. Rep. **52**, 263 (1979).
2. A. Lichtenberg and M. Lieberman, *Regular and Chaotic Dynamics* (Springer, Berlin, 1992).
3. G. M. Zaslavskiĭ and R. Z. Sagdeev, *Introduction to Nonlinear Physics* (Nauka, Moscow, 1988) [in Russian].
4. V. G. Gel'freĭkh and V. F. Lazutkin, Usp. Mat. Nauk **56** (3), 339 (2001).
5. S. Bullett, Commun. Math. Phys. **107**, 241 (1986).
6. V. V. Vecheslavov, Preprint No. 2000-27, IYaF SO RAN (Inst. of Nuclear Physics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 2000); nlin.CD/0005048.
7. V. V. Vecheslavov and B. V. Chirikov, Zh. Éksp. Teor. Fiz. **120**, 740 (2001) [JETP **93**, 649 (2001)].
8. V. V. Vecheslavov, Preprint No. 2003-11, IYaF SO RAN (Inst. of Nuclear Physics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 2003).
9. V. V. Vecheslavov, Zh. Tekh. Fiz. **58** (1), 20 (1988) [Sov. Phys. Tech. Phys. **33**, 11 (1988)].
10. V. V. Vecheslavov, Zh. Tekh. Fiz. **72** (2), 20 (2002) [Tech. Phys. **47**, 160 (2002)].
11. V. V. Vecheslavov, Zh. Éksp. Teor. Fiz. **125**, 399 (2004) [JETP **98**, 352 (2004)].
12. V. V. Vecheslavov, Zh. Éksp. Teor. Fiz. **109**, 2208 (1996) [JETP **82**, 1190 (1996)].
13. V. V. Vecheslavov, Pis'ma Zh. Éksp. Teor. Fiz. **63**, 989 (1996) [JETP Lett. **63**, 1047 (1996)].
14. V. V. Vecheslavov, Zh. Tekh. Fiz. **73** (9), 1 (2003) [Tech. Phys. **48**, 1079 (2003)].
15. G. M. Zaslavskiĭ and N. N. Filonenko, Zh. Éksp. Teor. Fiz. **54**, 1590 (1965) [Sov. Phys. JETP **27**, 851 (1965)].

Translated by A. Betev