Journal of Experimental and Theoretical Physics, Vol. 100, No. 1, 2005, pp. 1–13.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 1, 2005, pp. 5–17.

ATOMS, MOLECULES, OPTICS

Coulomb Corrections to Bremsstrahlung in the Electric Field of a Heavy Atom at High Energies¶

R. N. Lee[a], A. I. Milstein[a], V. M. Strakhovenko[a], and O. Ya. Schwarz[b]

[a]*Budker Institute of Nuclear Physics, Novosibirsk, 630090 Russia*
[b]*Novosibirsk State University, Novosibirsk, 630090 Russia*
e-mail: R.N.Lee@inp.nsk.su, A.I.Milstein@inp.nsk.su, V.M.Strakhovenko@inp.nsk.su

Received April 29, 2004

Abstract—We consider the differential and partially integrated cross sections for bremsstrahlung from high-energy electrons in an atomic field, with this field taken into account exactly. We use the semiclassical electron Green function and wavefunctions in an external electric field. It is shown that the Coulomb corrections to the differential cross section are very susceptible to screening. Nevertheless, the Coulomb corrections to the cross section summed over the final-electron states are independent of screening in the leading approximation in the small parameter $1/mr_{scr}$ (r_{scr} is the screening radius and m is the electron mass, $\hbar = c = 1$). We also consider bremsstrahlung from a finite-size electron beam on a heavy nucleus. The Coulomb corrections to the differential probability are also very susceptible to the beam shape, while the corrections to the probability integrated over momentum transfer are independent of it, apart from the trivial factor, which is the electron-beam density at zero impact parameter. For the Coulomb corrections to the bremsstrahlung spectrum, the next-to-leading terms with respect to the parameters m/ε (ε is the electron energy) and $1/mr_{scr}$ are obtained. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Bremsstrahlung in the electric field of atoms is a fundamental QED process. Its investigation, started in the 1930s, is important for various applications. In the Born approximation, both the differential cross section and the bremsstrahlung spectrum have been obtained for arbitrary electron energies and atomic form factors [1] (see also [2]). High-energy asymptotics of the bremsstrahlung cross section in a Coulomb field has been studied in detail in [3] exactly in the parameter $Z\alpha$ (where Z is the atomic number and $\alpha = 1/137$ is the fine-structure constant). In these papers, the differential cross sections and the bremsstrahlung spectrum have been obtained. For a screened Coulomb field, the high-energy asymptotics of the differential cross section was derived in [4]. The effect of screening on the spectrum was studied in [5, 6]. For the spectrum, it turned out that screening is essential only in the Born approximation. In other words, the Coulomb corrections to the spectrum are not significantly modified by screening. By definition, Coulomb corrections are the difference between the result obtained exactly in the external field and that obtained in the Born approximation. In the recent paper [7], it was claimed that Coulomb corrections to the differential cross section of the bremsstrahlung are also independent of screening.

In the present paper, we investigate the bremsstrahlung cross section in the electric field of a heavy atom. We assume that $\varepsilon, \varepsilon' \gg m$, where ε and ε' are the initial and final electron energies, respectively. In Section 2, we consider the differential cross section in detail in the leading approximation, i.e., neglecting corrections in the parameters m/ε and $1/mr_{scr}$. In contrast to the statement in [7], screening may strongly modify Coulomb corrections to the differential cross section. We demonstrate explicitly that this fact does not contradict the final-state integration theorem [5], which implies that Coulomb corrections to the spectrum are independent of screening. We also study the influence of the electron beam finite size on Coulomb corrections. Again, Coulomb corrections to the differential cross section are very sensitive to the shape of the electron beam, while the spectrum is independent of it, except for a trivial factor. In Section 3, we consider corrections to Coulomb corrections in the spectrum. It turns out that, in the first nonvanishing order, they enter the spectrum as a sum of two terms. The first term is proportional to m/ε and is independent of screening. The second term is small in the parameter $1/mr_{scr}$ and is independent of the energy.

Our approach is based on the use of the semiclassical Green function and the semiclassical wavefunction of the electron in an external field. Previously, this method was successfully applied to the investigation of the photoproduction process at high energy [8, 9].

¶ This article was submitted by authors in English.

2. DIFFERENTIAL CROSS SECTION

The cross section of the electron bremsstrahlung in the external field has the form

$$d\sigma^{\gamma} = \frac{\alpha}{(2\pi)^4\omega} d\mathbf{p}' d\mathbf{k}\,\delta(\varepsilon - \varepsilon' - \omega)|M|^2, \qquad (1)$$

where $\mathbf{k}$ is the photon momentum, $\mathbf{p}$ and $\mathbf{p}'$ are the respective initial and final electron momenta, and

$$\omega = |\mathbf{k}|, \quad \varepsilon = \varepsilon_p = \sqrt{\mathbf{p}^2 + m^2}, \quad \varepsilon' = \varepsilon_{p'}.$$

The matrix element M is given by

$$M = \int d\mathbf{r} \exp(-i\mathbf{k}\cdot\mathbf{r})\bar{\psi}^{(\mathrm{out})}_{P'}(\mathbf{r})\hat{e}^*\psi^{(\mathrm{in})}_{P}(\mathbf{r}), \qquad (2)$$

where $\psi^{(\mathrm{in})}_P$ and $\psi^{(\mathrm{out})}_P$ are the respective wavefunctions of the in- and out-state of the electron in an external field, containing the diverging and converging spherical waves and the plain wave with 4-momentum P in their asymptotics, $\hat{e}^* = e^*_\mu\gamma^\mu$, e_μ is the photon polarization 4-vector, and γ^μ are the Dirac matrices.

In [10], the semiclassical wavefunction of the electron in an arbitrary localized potential was found with the first correction in m/ε taken into account. In calculating bremsstrahlung and the e^+e^- photoproduction cross section in the leading approximation, the following form of the wavefunction can be used [10]:

$$\psi^{(\mathrm{in,\,out})}_P(\mathbf{r}) = \pm\int\frac{d\mathbf{q}}{i\pi}\exp\left[i\mathbf{p}\cdot\mathbf{r} \pm iq^2 \mp i\lambda\int_0^\infty dx V(\mathbf{r}_x)\right]$$
$$\times\left\{1 \mp \frac{1}{2p}\int_0^\infty dx\boldsymbol{\alpha}\cdot\nabla V(\mathbf{r}_x)\right\}u_P, \qquad (3)$$
$$\mathbf{r}_x = \mathbf{r} \mp x\mathbf{n} + \mathbf{q}\sqrt{2|\mathbf{r}\cdot\mathbf{n}|/p},$$
$$\lambda = \mathrm{sgn}P^0, \quad \mathbf{n} = \mathbf{p}/p.$$

In this formula, $\mathbf{q}$ is a two-dimensional vector lying in the plane perpendicular to $\mathbf{p}$, the upper sign corresponds to $\psi^{(\mathrm{in})}_P$, and u_P is the conventional Dirac spinor. We recall that the wavefunction $\psi^{(\mathrm{in})}_{(-\varepsilon_p,-\mathbf{p})}$ corresponds to the positron in the final state with the 4-momentum $(\varepsilon_p, \mathbf{p})$. For a Coulomb field, wavefunction (3) coincides with the standard Furry–Sommerfeld–Maue wavefunction. When the angles between $\mathbf{p}$ and $\mathbf{r}$ in $\psi^{(\mathrm{in})}_P(\mathbf{r})$ and between $\mathbf{p}$ and $-\mathbf{r}$ in $\psi^{(\mathrm{out})}_P(\mathbf{r})$ are not small, it is possible to replace $\mathbf{r}_x$ in Eq. (3) by $\mathbf{R}_x = \mathbf{r} \mp x\mathbf{n}$. Then, the integral over $\mathbf{q}$ can be taken, and we obtain the conventional eikonal wavefunction

$$\psi^{(\mathrm{in,\,out})}_{P,\,eik}(\mathbf{r}) = \exp\left[i\mathbf{p}\cdot\mathbf{r} \mp i\lambda\int_0^\infty dx V(\mathbf{R}_x)\right]$$
$$\times\left\{1 \mp \frac{1}{2p}\int_0^\infty dx\boldsymbol{\alpha}\cdot\nabla V(\mathbf{R}_x)\right\}u_P. \qquad (4)$$

We direct the z axis along the vector $\boldsymbol{\nu} = \mathbf{k}/\omega$, then $\mathbf{r} = z\boldsymbol{\nu} + \boldsymbol{\rho}$. In this frame, the polar angles of $\mathbf{p}$ and $\mathbf{p}'$ are small. We split the integration region in Eq. (2) in two: $z > 0$ and $z < 0$. The corresponding contributions to M are denoted as M_+ and M_-, with $M = M_+ + M_-$. For $z > 0$, the function $\psi^{(\mathrm{out})}_{p'}(\mathbf{r})$ has the eikonal form, and we obtain

$$M_+ = \int_{z>0} d\mathbf{r}\int\frac{d\mathbf{q}}{i\pi}\exp\left\{i\mathbf{q}^2 - i\boldsymbol{\Delta}\cdot\mathbf{r}\right.$$
$$\left. - i\int_0^\infty dx[V(\mathbf{r} - \mathbf{n}x + \mathbf{q}\sqrt{2z/p}) + V(\mathbf{r} + \mathbf{n}'x)]\right\}$$
$$\times\bar{u}_{p'}\left[\hat{e}^* - \frac{1}{2p}\int_0^\infty dx\hat{e}^*\boldsymbol{\alpha}\cdot\nabla V(\mathbf{r} - \mathbf{n}x + \mathbf{q}\sqrt{2z/p})\right.$$
$$\left. - \frac{1}{2p'}\int_0^\infty dx\boldsymbol{\alpha}\cdot\nabla V(\mathbf{r} + \mathbf{n}'x)\hat{e}^*\right]u_p, \qquad (5)$$

where $\boldsymbol{\Delta} = \mathbf{p}' + \mathbf{k} - \mathbf{p}$ is the momentum transfer.

In Eq. (5), we have replaced $\sqrt{2|\mathbf{r}\cdot\mathbf{n}|/p}$ in the definition of $\mathbf{r}_x$ in Eq. (3) by $\sqrt{2z/p}$. It is easy to see that, within our accuracy, we can also replace the quantity $V(\mathbf{r} + \mathbf{n}'x)$ in Eq. (5) by $V(\mathbf{r} + \mathbf{n}'x + \mathbf{q}\sqrt{2z/p})$ and consider the vector $\mathbf{q}$ to be perpendicular to z axis. After that, we shift $\boldsymbol{\rho} \longrightarrow \boldsymbol{\rho} - \mathbf{q}\sqrt{2z/p}$ and take the integral over $\mathbf{q}$. We obtain

$$M_+ = \int_{z>0} d\mathbf{r}\exp\left\{-i\frac{z}{2p}\Delta_\perp^2 - i\boldsymbol{\Delta}\cdot\mathbf{r}\right.$$
$$\left. - i\int_0^\infty dx[V(\mathbf{r} - \mathbf{n}x) + V(\mathbf{r} + \mathbf{n}'x)]\right\}$$
$$\times\bar{u}_{p'}\left[\hat{e}^* - \frac{1}{2p}\int_0^\infty dx\hat{e}^*\boldsymbol{\alpha}\cdot\nabla V(\mathbf{r} - \mathbf{n}x)\right.$$
$$\left. - \frac{1}{2p'}\int_0^\infty dx\boldsymbol{\alpha}\cdot\nabla V(\mathbf{r} + \mathbf{n}'x)\hat{e}^*\right]u_p. \qquad (6)$$

In the same way, we obtain

$$M_- = \int\limits_{z<0} d\mathbf{r}\exp\left\{i\frac{z}{2p'}\Delta_\perp^2 - i\boldsymbol{\Delta}\cdot\mathbf{r}\right.$$
$$\left. - i\int\limits_0^\infty dx[V(\mathbf{r}-\mathbf{n}x)+V(\mathbf{r}+\mathbf{n}'x)]\right\}$$
$$\times\bar{u}_{p'}\left[\hat{e}^* - \frac{1}{2p}\int\limits_0^\infty dx\hat{e}^*\boldsymbol{\alpha}\cdot\nabla V(\mathbf{r}-\mathbf{n}x)\right.$$
$$\left. - \frac{1}{2p'}\int\limits_0^\infty dx\boldsymbol{\alpha}\cdot\nabla V(\mathbf{r}+\mathbf{n}'x)\hat{e}^*\right]u_p. \quad (7)$$

There are two overlapping regions of the momentum transfer Δ,

$$\text{I. } \Delta \ll \frac{m\omega}{\varepsilon},$$
$$\text{II. } \Delta \gg \Delta_{\min} = \frac{m^2\omega}{2\varepsilon\varepsilon'}. \quad (8)$$

In the first region, we can neglect the terms proportional to $\Delta_\perp^2$ in the exponents in Eqs. (6) and (7). Then, the sum

$$M = M_+ + M_-$$

becomes

$$M = \int d\mathbf{r}\exp\left\{-i\boldsymbol{\Delta}\cdot\mathbf{r}\right.$$
$$\left. - i\int\limits_0^\infty dx[V(\mathbf{r}-\mathbf{n}x)+V(\mathbf{r}+\mathbf{n}'x)]\right\}$$
$$\times\bar{u}_{p'}\left[\hat{e}^* - \frac{1}{2p}\int\limits_0^\infty dx\hat{e}^*\boldsymbol{\alpha}\cdot\nabla V(\mathbf{r}-\mathbf{n}x)\right.$$
$$\left. - \frac{1}{2p'}\int\limits_0^\infty dx\boldsymbol{\alpha}\cdot\nabla V(\mathbf{r}+\mathbf{n}'x)\hat{e}^*\right]u_p. \quad (9)$$

We can make the replacement $\mathbf{n}, \mathbf{n}' \longrightarrow \boldsymbol{\nu}$ in the prefactor in Eq. (9). In the exponent, we must take the linear term of the expansion of the integral in $\mathbf{n}-\boldsymbol{\nu}$ and $\mathbf{n}'-\boldsymbol{\nu}$ into account. As a result, we have

$$M = \int d\mathbf{r}\exp[-i\boldsymbol{\Delta}\cdot\mathbf{r} - i\chi(\boldsymbol{\rho})]$$
$$\times\int\limits_0^\infty dy\bar{u}_{p'}[\hat{e}^*[iy(\mathbf{n}-\boldsymbol{\nu}) - \boldsymbol{\alpha}/2p]\cdot\nabla V(\mathbf{r}-\boldsymbol{\nu}y)$$
$$+ [-iy(\mathbf{n}'-\boldsymbol{\nu}) - \boldsymbol{\alpha}/2p']\cdot\nabla V(\mathbf{r}+\boldsymbol{\nu}y)\hat{e}^*]u_p, \quad (10)$$
$$\chi(\boldsymbol{\rho}) = \int\limits_{-\infty}^\infty dz V(\mathbf{r}).$$

In the arguments of $V(\mathbf{r}\pm\boldsymbol{\nu}y)$, we make the substitutions $z \longrightarrow z \mp y$. After that, we take the integral over y and obtain

$$M = \mathbf{A}(\boldsymbol{\Delta})\cdot\left(\bar{u}_{p'}\left[\frac{(\mathbf{n}-\mathbf{n}')\hat{e}^*}{\Delta_z^2} - \frac{\hat{e}^*\boldsymbol{\alpha}}{2p\Delta_z} + \frac{\boldsymbol{\alpha}\hat{e}^*}{2p'\Delta_z}\right]u_p\right),$$
$$\mathbf{A}(\boldsymbol{\Delta}) = -i\int d\mathbf{r}\exp[-i\boldsymbol{\Delta}\cdot\mathbf{r} - i\chi(\boldsymbol{\rho})]\nabla_\rho V(\mathbf{r}). \quad (11)$$

We now pass to the calculation of M in the second region, where $\Delta \gg \Delta_{\min}$. In Eq. (6) for M_+, we can replace $\mathbf{n}' \longrightarrow \mathbf{n}$ and $z\Delta_\perp^2/2p \longrightarrow \tilde{z}\Delta_\perp^2/2p$, where $\tilde{z} = \mathbf{r}\cdot\mathbf{n}$. Because the polar angle of $\mathbf{n}$ is small, we can integrate in Eq. (6) over the half-space $\tilde{z} > 0$. After the integration over $\tilde{z}$, we obtain

$$M_+ = -i\int d\boldsymbol{\rho}\exp[-i\boldsymbol{\Delta}\cdot\boldsymbol{\rho} - i\chi(\boldsymbol{\rho})]$$
$$\times\frac{\bar{u}_{p'}\hat{e}^*[2p + \boldsymbol{\alpha}\cdot\boldsymbol{\Delta}_\perp]u_p}{2p\boldsymbol{\Delta}\cdot\mathbf{n} + \Delta_\perp^2}. \quad (12)$$

The calculation of M_- is performed quite similarly. As a result, we have

$$M = -i\int d\boldsymbol{\rho}\exp[-i\boldsymbol{\Delta}\cdot\boldsymbol{\rho} - i\chi(\boldsymbol{\rho})]$$
$$\times\bar{u}_{p'}\left[\frac{\hat{e}^*(2p + \boldsymbol{\alpha}\cdot\boldsymbol{\Delta}_\perp)}{2p\boldsymbol{\Delta}\cdot\mathbf{n} + \Delta_\perp^2} - \frac{(2p' + \boldsymbol{\alpha}\cdot\boldsymbol{\Delta}_\perp)\hat{e}^*}{2p'\boldsymbol{\Delta}\cdot\mathbf{n}' - \Delta_\perp^2}\right]u_p. \quad (13)$$

Now, we can write the representation for M that is valid in both regions,

$$M = \frac{\varepsilon\varepsilon'}{\omega}\mathbf{A}(\boldsymbol{\Delta})\cdot\left\{\bar{u}_{p'}\left[-2\hat{e}^*\frac{\mathbf{p}_\perp + \mathbf{p}'_\perp}{\delta\delta'} + \frac{\hat{e}^*\boldsymbol{\alpha}}{\varepsilon\delta'} - \frac{\boldsymbol{\alpha}\hat{e}^*}{\varepsilon'\delta}\right]u_p\right\},$$
$$\delta = m^2 + \mathbf{p}_\perp^2, \quad \delta' = m^2 + \mathbf{p}'^2_\perp. \quad (14)$$

Within our accuracy, this expression coincides with Eq. (11) in region I and with Eq. (13) in region II. Using the explicit form of the Dirac spinors, we finally obtain

$$M = \frac{1}{2\delta\delta'}\mathbf{A}(\boldsymbol{\Delta})\cdot\left\{\varphi'^{\dagger}\left[(\mathbf{p}_\perp+\mathbf{p}'_\perp)\right.\right.$$
$$\times\left(\frac{\varepsilon+\varepsilon'}{\omega}\mathbf{e}^*\cdot(\mathbf{p}_\perp+\mathbf{p}'_\perp)-i[\boldsymbol{\sigma}\times\mathbf{e}^*]\cdot(\mathbf{p}_\perp+\mathbf{p}'_\perp)\right.$$
$$\left.+2im[\boldsymbol{\sigma}\times\mathbf{e}^*]_z\right) \tag{15}$$
$$\left.\left.-(\delta+\delta')\left(\frac{\varepsilon+\varepsilon'}{\omega}\mathbf{e}^*-i[\boldsymbol{\sigma}\times\mathbf{e}^*]_\perp\right)\right]\varphi\right\}.$$

This expression is in agreement with that obtained in [4] by another method. We emphasize that the potential enters amplitude (15) only via $\mathbf{A}(\boldsymbol{\Delta})$.

2.1. Coulomb Corrections to the Differential Cross Section in a Screened Coulomb Potential

We discuss Coulomb corrections to the differential cross section of bremsstrahlung. We recall that these corrections are the difference between the exact (in the external field strength) cross section and that obtained in the Born approximation, which is proportional to $[|\mathbf{A}(\boldsymbol{\Delta})|^2-|\mathbf{A}_B(\boldsymbol{\Delta})|^2]$ with $\mathbf{A}(\boldsymbol{\Delta})$ from Eq. (11) and

$$\mathbf{A}_B(\boldsymbol{\Delta}) = -i\int d\mathbf{r}\exp[-i\boldsymbol{\Delta}\cdot\mathbf{r}]\nabla_\rho V(\mathbf{r})$$
$$= \boldsymbol{\Delta}_\perp\int d\mathbf{r}\exp[-i\boldsymbol{\Delta}\cdot\mathbf{r}]V(\mathbf{r}). \tag{16}$$

The screening modifies the Coulomb potential of the nucleus at distances

$$r_{scr}\gg\lambda_C = 1/m.$$

In the region

$$\Delta\gg\max(\Delta_{\min}, r_{scr}^{-1}),$$

the quantities $\mathbf{A}(\boldsymbol{\Delta})$ and $\mathbf{A}_B(\boldsymbol{\Delta})$ are of the form

$$\mathbf{A}(\boldsymbol{\Delta}) = \mathbf{A}_B(\boldsymbol{\Delta})\frac{\Gamma(1-iZ\alpha)}{\Gamma(1+iZ\alpha)}\left(\frac{4}{\Delta_\perp^2}\right)^{-iZ\alpha}$$
$$= -\boldsymbol{\Delta}_\perp\pi Z\alpha\frac{\Gamma(1-iZ\alpha)}{\Gamma(1+iZ\alpha)}\left(\frac{4}{\Delta_\perp^2}\right)^{1-iZ\alpha}. \tag{17}$$

Therefore,

$$|A(\boldsymbol{\Delta})|^2 = |A_B(\boldsymbol{\Delta})|^2 \text{ for } \Delta\gg\max(\Delta_{\min}, r_{scr}^{-1})$$

and Coulomb corrections to the differential cross section vanish in this region in the leading approximation. Hence, Coulomb corrections are important only in the region

$$\Delta\lesssim\max(\Delta_{\min}, r_{scr}^{-1})\ll m.$$

In this region, we can use Eq. (11) for the matrix element. For the Coulomb corrections, substituting Eq. (11) in Eq. (1), using the relation

$$d\Omega_{\mathbf{p}'}d\Omega_{\mathbf{k}} = \frac{d\phi d\boldsymbol{\Delta}_\perp d\Delta_z}{\omega\varepsilon\varepsilon'},$$

and integrating over the azimuthal angle ϕ and summating over polarizations, we obtain

$$d\sigma_C^\gamma = \frac{\alpha d\omega d\boldsymbol{\Delta}_\perp d\Delta_z}{16\pi^3\varepsilon^3\varepsilon'\Delta_z^2}$$
$$\times\left[\varepsilon^2+\varepsilon'^2+2\frac{m^2\omega}{\Delta_z}+\frac{m^4\omega^2}{\varepsilon\varepsilon'\Delta_z^2}\right][|\mathbf{A}(\boldsymbol{\Delta})|^2-|\mathbf{A}_B(\boldsymbol{\Delta})|^2]. \tag{18}$$

We note that, in this formula, we can assume that the z axis is directed along the vector $\mathbf{p}$. Then, Δ_z is negative and

$$|\Delta_z|\geq\Delta_{\min} = \frac{m^2\omega}{2\varepsilon\varepsilon'}.$$

The potential $V(\mathbf{r})$ and the transverse momentum transfer $\boldsymbol{\Delta}_\perp$ enter Eq. (18) only as the factor dR,

$$dR = d\boldsymbol{\Delta}_\perp[|\mathbf{A}(\boldsymbol{\Delta})|^2-|\mathbf{A}_B(\boldsymbol{\Delta})|^2]. \tag{19}$$

It follows from the definition of $\mathbf{A}(\boldsymbol{\Delta})$ that, for $r_{scr}\gg|\Delta_z|^{-1}$, screening can be neglected. However, it is obvious from Eq. (19) that screening drastically modifies the $\boldsymbol{\Delta}_\perp$-dependence of the differential cross section for $r_{scr}\lesssim|\Delta_z|^{-1}$. We illustrate this statement with the example of the Yukawa potential

$$V(r) = -Z\alpha\exp[-\beta r]/r.$$

After straightforward calculation, we have

$$\Delta_\perp\frac{dR}{d\Delta_\perp} = 32\pi^3(Z\alpha)^2$$
$$\times\left[\zeta^2\left|\int_0^\infty dx x J_1(x\zeta)K_1(x)\exp[2iZ\alpha K_0(\gamma x)]\right|^2\right.$$
$$\left.-\frac{\zeta^4}{(1+\zeta^2)^2}\right], \tag{20}$$
$$\zeta = \frac{\Delta_\perp}{\sqrt{\Delta_z^2+\beta^2}},\quad \gamma = \frac{\beta}{\sqrt{\Delta_z^2+\beta^2}}.$$

We emphasize that $\Delta_\perp$ enters the right-hand side of Eq. (20) only via the variable ζ, and, hence, $\sqrt{\Delta_z^2 + \beta^2}$ is the characteristic scale of distribution (20). For $\beta \gg |\Delta_z|$, this scale is entirely determined by the screening radius $r_{scr} = \beta^{-1}$. In this case, the $\Delta_\perp$-distribution is much wider than that in the absence of screening. We therefore conclude that, in contrast to the statement in [7], Coulomb corrections to the differential cross section strongly depend on screening. We note that screening also affects the shape of the $\Delta_\perp$-distribution (20) via the parameter γ, which varies from 0 to 1. In Fig. 1, we show the dependence of $\Delta_\perp dR/d\Delta_\perp$ on the scaling variable ζ for $Z = 80$ and different values of the parameter γ.

We note that, in contrast to bremsstrahlung, Coulomb corrections to the differential cross section of e^+e^- photoproduction in the atomic field are important only in the region $\Delta_\perp \sim m$, where screening can be neglected [4].

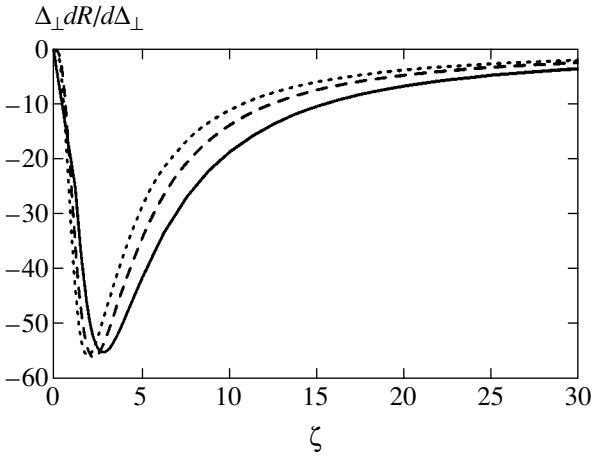


Fig. 1. The quantity $\Delta_\perp dR/d\Delta_\perp$ as a function of ζ for $Z = 80$ and $\gamma = 1$ (solid curve), $\gamma = 0.5$ (dashed curve), and $\gamma = 0.01$ (dotted curve). The variable ζ is defined in Eq. (20).

2.2. Integrated Cross Section

It was shown in [5] that Coulomb corrections to the cross section of bremsstrahlung integrated over $\boldsymbol{\Delta}_\perp$ are independent of screening in the leading approximation. The statement was based on the possibility of obtaining this cross section from the one for the e^+e^- photoproduction. In this subsection, we perform the explicit integration of $d\sigma_C^\gamma$ (Eq. (18)) over $\boldsymbol{\Delta}_\perp$. We show that the strong influence of screening on the shape of $d\sigma_C^\gamma$ does not contradict the statement in [5]. Our consideration is quite similar to that used in [11] in the calculation of Coulomb corrections to the e^+e^- pair production in ultrarelativistic heavy-ion collisions.

We consider the quantity

$$R = \int dR = \int d\boldsymbol{\Delta}_\perp [|\mathbf{A}(\boldsymbol{\Delta})|^2 - |\mathbf{A}_B(\boldsymbol{\Delta})|^2]. \qquad (21)$$

This integral converges due to the compensation in the integrand, and the main contribution comes from the region

$$\Delta_\perp \lesssim \max(|\Delta_z|, r_{scr}^{-1})$$

(see Eq. (17)). Substituting the integral representation for $\mathbf{A}(\boldsymbol{\Delta})$ (Eq. (11)) and for $\mathbf{A}_B(\boldsymbol{\Delta})$ (Eq. (16)) in Eq. (21), we have

$$R = \int d\boldsymbol{\Delta}_\perp \iint d\mathbf{r}_1 d\mathbf{r}_2 \times \exp[i\boldsymbol{\Delta}\cdot(\mathbf{r}_1 - \mathbf{r}_2)]\{\exp[i\chi(\boldsymbol{\rho}_1) - i\chi(\boldsymbol{\rho}_2)] - 1\} \times [\nabla_{1\perp} V(\mathbf{r}_1)]\cdot[\nabla_{2\perp} V(\mathbf{r}_2)]. \qquad (22)$$

It is necessary to treat this repeated integral with care. If one naively changes the order of integration over $\boldsymbol{\Delta}_\perp$ and $\mathbf{r}_{1,2}$, the integration over $\boldsymbol{\Delta}_\perp$ in infinite limits leads to $\delta(\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2)$. Then, the quantity R vanishes after the integration over $\boldsymbol{\rho}_1$, which is not correct. Such an erroneous change of the order of integrations was made in [4] in explicitly verifying that the integrated cross section is independent of screening. Although this independence itself takes place, the proof of this fact given in [4] and widely cited in textbooks is not consistent. The correct integration in Eq. (22) can be performed as follows. We restrict the region of integration over $\boldsymbol{\Delta}_\perp$ by the condition

$$\Delta_\perp < Q,$$

where

$$Q \gg \max(|\Delta_z|, r_{scr}^{-1}).$$

In this region, integral (21) is saturated and, hence, the result of integration must be independent of Q. We can then change the order of integrations over $\mathbf{r}_{1,2}$ and $\boldsymbol{\Delta}_\perp$ in Eq. (22) and take the integral over $\boldsymbol{\Delta}_\perp$:

$$R = 2\pi Q \iint d\mathbf{r}_1 d\mathbf{r}_2 \frac{J_1(Q|\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2|)}{|\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2|} \exp[i\Delta_z(z_1 - z_2)] \times \{\exp[i\chi(\boldsymbol{\rho}_1) - i\chi(\boldsymbol{\rho}_2)] - 1\}[\nabla_{1\perp} V(\mathbf{r}_1)]\cdot[\nabla_{2\perp} V(\mathbf{r}_2)]. \qquad (23)$$

It is seen from this formula that the main contribution to the integral is given by the region $|\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2| \sim 1/Q$. If $\rho_{1,2} \gg 1/Q$ and $|\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2| \sim 1/Q$; then

$$|\exp[i\chi(\boldsymbol{\rho}_1) - i\chi(\boldsymbol{\rho}_2)] - 1| \ll 1,$$

and the integrand is suppressed. Therefore, integral (23) is determined by the region where both $\rho_1 \sim 1/Q$ and

$\rho_2 \sim 1/Q$. Due to the factor $\nabla_{1\perp}V(\mathbf{r}_1)\nabla_{2\perp}V(\mathbf{r}_2)$ in the integrand, $z_{1,2} \sim 1/Q$ also. If $r \ll r_{scr}$, then

$$V(\mathbf{r}) \approx -Z\alpha/r$$

and

$$\chi(\boldsymbol{\rho}) \approx 2Z\alpha(\ln\rho + \text{const}).$$

In addition, for $r_{1,2} \ll |\Delta_z|^{-1}$, we can omit the factor $\exp[i\Delta_z(z_1 - z_2)]$ in (23). We then perform the substitution $\mathbf{r}_{1,2} \longrightarrow \mathbf{r}_{1,2}/Q$ and obtain

$$R = 8\pi(Z\alpha)^2\int\int d\boldsymbol{\rho}_1 d\boldsymbol{\rho}_2 \times \frac{(\boldsymbol{\rho}_1\cdot\boldsymbol{\rho}_2)J_1(|\boldsymbol{\rho}_1-\boldsymbol{\rho}_2|)}{\rho_1^2\rho_2^2|\boldsymbol{\rho}_1-\boldsymbol{\rho}_2|}\left\{\left(\frac{\rho_2}{\rho_1}\right)^{2iZ\alpha} - 1\right\}. \quad (24)$$

We emphasize that this formula does not contain Q. Using the identity

$$\frac{(\boldsymbol{\rho}_1\cdot\boldsymbol{\rho}_2)J_1(|\boldsymbol{\rho}_1-\boldsymbol{\rho}_2|)}{|\boldsymbol{\rho}_1-\boldsymbol{\rho}_2|} = \frac{\rho_1\rho_2}{\rho_1^2-\rho_2^2} \times\left(\rho_1\frac{\partial}{\partial\rho_2} - \rho_2\frac{\partial}{\partial\rho_1}\right)J_0(|\boldsymbol{\rho}_1-\boldsymbol{\rho}_2|), \quad (25)$$

and the relation

$$\int_0^{2\pi} d\phi J_0(\sqrt{\rho_1^2+\rho_2^2-2\rho_1\rho_2\cos\phi}) = 2\pi J_0(\rho_1)J_0(\rho_2), \quad (26)$$

which follows from the summation theorem for the Bessel functions, we have

$$R = 32\pi^3(Z\alpha)^2\int_0^\infty\int_0^\infty \frac{d\rho_1 d\rho_2}{\rho_1^2-\rho_2^2} \times[\rho_2 J_0(\rho_2)J_1(\rho_1) - \rho_1 J_0(\rho_1)J_1(\rho_2)]\left\{\left(\frac{\rho_2}{\rho_1}\right)^{2iZ\alpha} - 1\right\}. \quad (27)$$

Making the change of variables $\rho_{1,2} = r\exp(\pm t/4)$ and integrating over r, we finally obtain

$$R = 32\pi^3(Z\alpha)^2\int_0^\infty dt\frac{\cos(Z\alpha t) - 1}{\exp(t) - 1} = -32\pi^3(Z\alpha)^2[\mathrm{Re}\,\psi(1 + iZ\alpha) + C] = -32\pi^3(Z\alpha)^2 f(Z\alpha), \quad (28)$$

where C is the Euler constant and

$$\psi(x) = d\ln\Gamma(x)/dx.$$

Using this formula and taking the integral over Δ_z from $-\infty$ to $-\Delta_{\min}$ in Eq. (18), we reproduce the well-known result obtained in [3]. We note that the value of R following from the numerical integration of Eq. (20) over $\Delta_\perp$ agrees with the universal result (28).

Thus, we come to a remarkable conclusion: Coulomb corrections to the cross section integrated over $\boldsymbol{\Delta}_\perp$ are independent of screening, although the main contribution to the integral comes from the region

$$\Delta_\perp \lesssim \max(\Delta_{\min}, r_{scr}^{-1}),$$

where, for $\Delta_{\min} \ll r_{scr}^{-1}$, the differential cross section is essentially modified by screening. We emphasize that this result is valid in the leading approximation with respect to the parameters $m/\varepsilon \ll 1$ and $\lambda_C/r_{scr} \ll 1$. In the next section, we show that, in the limit $m/\varepsilon \longrightarrow 0$, the screening contributes to $d\sigma_C^\gamma/d\omega$ only as a correction in the parameter λ_C/r_{scr}.

2.3. Effect of Beam Size on Coulomb Corrections

It is interesting to consider the effect of a finite transverse size b of the electron beam on Coulomb corrections to bremsstrahlung in a Coulomb field of a heavy nucleus. This consideration should be performed in terms of the probability dW rather than the cross section. Similarly to the effect of screening, a finite beam size can lead to a substantial modification of Coulomb corrections to the differential probability dW_C, while Coulomb corrections to the probability integrated over $\boldsymbol{\Delta}$ is a universal function. To illustrate this statement, we consider bremsstrahlung from the electron described in the initial state by the wavefunction of the form

$$\psi(\mathbf{r}) = \int d\Omega_{\mathbf{p}} h(\mathbf{p})\psi_P^{(\mathrm{in})}(\mathbf{r}), \quad (29)$$

where the function $h(\mathbf{p})$ peaks at $\mathbf{p} = \mathbf{p}_0$. If the width δp of the peak satisfies the condition

$$\delta p \ll \sqrt{\Delta_{\min}\varepsilon} \lesssim m,$$

then

$$\psi(\mathbf{r}) \approx \int d\Omega_{\mathbf{p}} h(\mathbf{p})\exp[i(\mathbf{p}-\mathbf{p}_0)\cdot\boldsymbol{\rho}]\psi_{P_0}^{(\mathrm{in})}(\mathbf{r}) = \phi(\boldsymbol{\rho})\psi_{P_0}^{(\mathrm{in})}(\mathbf{r}), \quad (30)$$

where the function $\phi(\boldsymbol{\rho})$ is normalized as

$$\int d\boldsymbol{\rho}|\phi(\boldsymbol{\rho})|^2 = 1$$

and has a width of

$$b \gg 1/\sqrt{\Delta_{\min}\varepsilon} \gtrsim \lambda_C.$$

The quantity dW_C is given by the right-hand side of formula (18), where the functions $\mathbf{A}(\boldsymbol{\Delta})$ and $\mathbf{A}_B(\boldsymbol{\Delta})$ are given by Eqs. (11) and (16) with the additional factor $\phi(\boldsymbol{\rho})$ in the integrands. Substituting

$$V(r) = -Z\alpha/r,$$

we have

$$\begin{aligned}\mathbf{A}(\boldsymbol{\Delta}) &= -2iZ\alpha\Delta_z\int d\boldsymbol{\rho}\phi(\boldsymbol{\rho})\\ &\times \exp[-i\boldsymbol{\Delta}_\perp\cdot\boldsymbol{\rho}]K_1(\Delta_z\rho)\boldsymbol{\rho}/\rho^{1+2iZ\alpha},\\ \mathbf{A}_B(\boldsymbol{\Delta}) &= -2iZ\alpha\Delta_z\int d\boldsymbol{\rho}\phi(\boldsymbol{\rho})\\ &\times \exp[-i\boldsymbol{\Delta}_\perp\cdot\boldsymbol{\rho}]K_1(\Delta_z\rho)\boldsymbol{\rho}/\rho.\end{aligned} \tag{31}$$

If

$$b \gg |\Delta_z|^{-1} \sim \Delta_{\min}^{-1},$$

then we can simply replace $\phi(\boldsymbol{\rho}) \longrightarrow \phi(0)$ in Eq. (31), so that the differential distribution does not change compared with the case of a plain wave. Therefore, we consider the case $b \ll \Delta_{\min}^{-1}$, where the finiteness of the beam size is very important. In this case, we can replace $K_1(\Delta_z\rho) \longrightarrow (\Delta_z\rho)^{-1}$ in Eq. (31).

Substituting the functions $\mathbf{A}(\boldsymbol{\Delta}_\perp)$ and $\mathbf{A}_B(\boldsymbol{\Delta}_\perp)$ from Eq. (31) in dR defined by Eq. (19) and repeating all the steps of the derivation of

$$R = \int dR$$

in the previous subsection, we obtain

$$R = -32\pi^3(Z\alpha)^2 f(Z\alpha)|\phi(0)|^2. \tag{32}$$

We see that Coulomb corrections to the integrated probability depend on the shape of the wave packet only via the factor $|\phi(0)|^2$, corresponding to a electron density at zero impact parameter. Therefore, their dependence on $Z\alpha$ coincides with that in the case of a plane wave (24). However, the shape of $\phi(\boldsymbol{\rho})$ can essentially modify the $\boldsymbol{\Delta}_\perp$-dependence of dW_C. As an illustration, in Fig. 2, we show the dependence of $\Delta_\perp dR/d\Delta_\perp$ on ζ for $Z = 80$ and

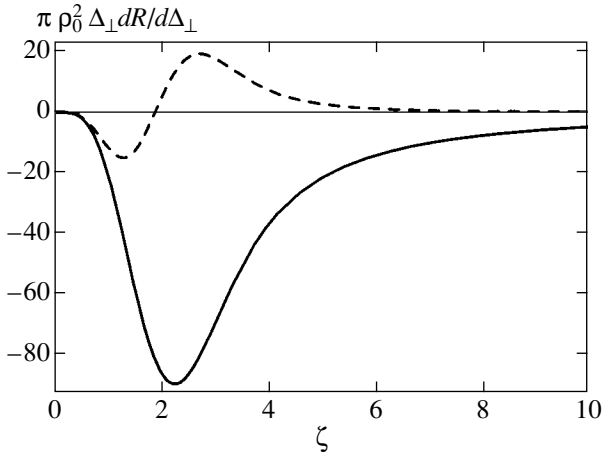


Fig. 2. The quantity $\Delta_\perp dR/d\Delta_\perp$ in the units $(\pi\rho_0^2)^{-1}$ as a function of $\zeta = \rho_0\Delta_\perp$ for $Z = 80$ and $\phi(\boldsymbol{\rho}) = \phi_0(\rho)$ (solid curve), $\phi(\boldsymbol{\rho}) = \phi_1(\rho)$ (dashed curve). The functions $\phi_{0,1}$ are defined in Eq. (33).

$\phi(\boldsymbol{\rho}) = \phi_0(\rho)$ (solid curve) and $\phi(\boldsymbol{\rho}) = \phi_1(\rho)$ (dashed curve), where

$$\begin{aligned}\phi_0(\boldsymbol{\rho}) &= \frac{\exp[-\rho^2/2\rho_0^2]}{\sqrt{\pi\rho_0^2}},\\ \phi_1(\boldsymbol{\rho}) &= \frac{(\rho/\rho_0)^2\exp[-\rho^2/2\rho_0^2]}{\sqrt{2\pi\rho_0^2}}, \quad \zeta = \rho_0\Delta_\perp.\end{aligned} \tag{33}$$

It is seen that the behavior of $\Delta_\perp dR/d\Delta_\perp$ differs drastically for the two cases considered. In accordance with Eq. (32),

$$R = -32\pi^3(Z\alpha)^2 f(Z\alpha)/\pi\rho_0^2$$

for $\phi(\boldsymbol{\rho}) = \phi_0(\rho)$ and

$$R = 0$$

for $\phi(\boldsymbol{\rho}) = \phi_1(\rho)$. We note that, in the latter case, the function $\Delta_\perp dR/d\Delta_\perp$ itself is different from zero.

3. NEXT-TO-LEADING TERMS IN THE BREMSSTRAHLUNG SPECTRUM

As is known [5], the modification of the high-energy asymptotics of Coulomb corrections to the spectrum due to the screening effect is small. Below, we show that the same is also true for the next term in m/ε. In this section, we explicitly calculate the screening correction in the leading term of the high-energy asymptotics and neglect screening in calculating the next-to-leading term in m/ε. In other words, we calculate the first cor-

rections in the small parameters m/ε and $1/mr_{scr}$ to the bremsstrahlung spectrum

$$\frac{d\sigma^{\gamma}}{d\omega} = \frac{\alpha\omega p'\varepsilon'}{2(2\pi)^4}\int d\Omega_{\mathbf{p}'}d\Omega_{\mathbf{k}}\sum_{\lambda_e\lambda_e',\lambda_\gamma}|M|^2, \tag{34}$$

with the amplitude M given by Eq. (2) and summation performed over the polarizations of all particles. It is convenient to calculate $d\sigma^{\gamma}/d\omega$ using the Green function $G(\mathbf{r}_2, \mathbf{r}_2|\varepsilon)$ of the Dirac equation in an external field. This Green function can be represented as

$$G(\mathbf{r}_2, \mathbf{r}_1|\varepsilon) = \sum_{\lambda_e, n}\frac{\psi_n(\mathbf{r}_2)\bar{\psi}_n(\mathbf{r}_1)}{\varepsilon-\varepsilon_n+i0} + \sum_{\lambda_e}\int\frac{d\mathbf{p}}{(2\pi)^3}\left[\frac{\psi_P(\mathbf{r}_2)\bar{\psi}_P(\mathbf{r}_1)}{\varepsilon-\varepsilon_p+i0}+\frac{\psi_{-P}(\mathbf{r}_2)\bar{\psi}_{-P}(\mathbf{r}_1)}{\varepsilon+\varepsilon_p-i0}\right], \tag{35}$$

where ψ_n is the discrete-spectrum wavefunction, ε_n is the corresponding binding energy, and $P = (\varepsilon_p, \mathbf{p})$. The set of either in- or out-wavefunctions can be used in Eq. (35). The regularization of denominators in Eq. (35) corresponds to the Feynman rule. From Eq. (35),

$$\sum_{\lambda_e}\int d\Omega_{\mathbf{p}}\psi_P^{(\mathrm{in})}(\mathbf{r}_1)\bar{\psi}_P^{(\mathrm{in})}(\mathbf{r}_2) = \sum_{\lambda_e}\int d\Omega_{\mathbf{p}}\psi_P^{(\mathrm{out})}(\mathbf{r}_1)\bar{\psi}_P^{(\mathrm{out})}(\mathbf{r}_2) = i\frac{(2\pi)^2}{\varepsilon_p p}\delta G(\mathbf{r}_1, \mathbf{r}_2|\varepsilon_p), \tag{36}$$

where $\Omega_{\mathbf{p}}$ is the solid angle of $\mathbf{p}$ and $\delta G = G - \tilde{G}$. The function $\tilde{G}$ is obtained from (35) by the replacement $i0 \longrightarrow -i0$. Because the bremsstrahlung spectrum is independent of the direction of the vector $\mathbf{p}$, we can average the right-hand side of Eq. (34) over the angles of this vector. Using Eq. (36), we then obtain

$$\frac{d\sigma^{\gamma}}{d\omega} = -\frac{\alpha\omega}{2\varepsilon p}\int\frac{d\Omega_k}{4\pi}\int\int d\mathbf{r}_1 d\mathbf{r}_2\exp(-i\mathbf{k}\cdot\mathbf{r}) \times\sum_{\lambda_\gamma}\mathrm{Sp}\{\delta G(\mathbf{r}_2, \mathbf{r}_1|\varepsilon)\hat{e}\delta G(\mathbf{r}_1, \mathbf{r}_2|\varepsilon')\hat{e}\}, \tag{37}$$

where $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$ and $\varepsilon' = \varepsilon - \omega$ is the energy of the final electron. Here and below, we use the linear polarization basis ($\mathbf{e}^* = \mathbf{e}$). We note that the integration over $d\Omega_{\mathbf{k}}$ is trivial because the integrand is independent of the angles of $\mathbf{k}$, and we therefore omit the integral $\int d\Omega_k/4\pi$ below. It is convenient to represent $d\sigma^{\gamma}/d\omega$ in another form using the Green function $D(\mathbf{r}_2, \mathbf{r}_1|\varepsilon)$ of the squared Dirac equation,

$$G(\mathbf{r}_2, \mathbf{r}_1|\varepsilon) = [\gamma^0(\varepsilon - V(\mathbf{r}_2)) - \boldsymbol{\gamma}\cdot\mathbf{p}_2 + m]D(\mathbf{r}_2, \mathbf{r}_1|\varepsilon), \quad \mathbf{p}_2 = -i\nabla_2. \tag{38}$$

Performing transformations as in [9, 12], we can rewrite Eq. (37) as

$$\frac{d\sigma^{\gamma}}{d\omega} = -\frac{\alpha\omega}{4\varepsilon p}\int\int d\mathbf{r}_1 d\mathbf{r}_2\exp(-i\mathbf{k}\cdot\mathbf{r}) \times\sum_{\lambda_\gamma}\mathrm{Sp}\{[(2\mathbf{e}\cdot\mathbf{p}_2 - \hat{e}\hat{k})\delta D(\mathbf{r}_2, \mathbf{r}_1|\varepsilon)] \times[(2\mathbf{e}\cdot\mathbf{p}_1 + \hat{e}\hat{k})\delta D(\mathbf{r}_1, \mathbf{r}_2|\varepsilon')]\}. \tag{39}$$

For the first two terms of the high-energy asymptotic expansion of the spectrum, the leading contribution to the integral in Eqs. (37) and (39) is given by the region

$$r = |\mathbf{r}_2 - \mathbf{r}_1| \sim \frac{1}{\Delta_{\min}} = \frac{2\varepsilon\varepsilon'}{\omega m^2} \gg \frac{1}{m}.$$

This estimate is in accordance with the uncertainty relation. Substituting $\delta D = D - \tilde{D}$ in Eq. (39), we obtain four terms. Within our accuracy, the terms containing $D(\varepsilon)D(\varepsilon')$ and $\tilde{D}(\varepsilon)\tilde{D}(\varepsilon')$ can be omitted, and we have

$$\frac{d\sigma^{\gamma}}{d\omega} = \frac{\alpha\omega}{2\varepsilon p}\mathrm{Re}\int\int d\mathbf{r}_1 d\mathbf{r}_2\exp(-i\mathbf{k}\cdot\mathbf{r}) \times\sum_{\lambda_\gamma}\mathrm{Sp}\{[(2\mathbf{e}\cdot\mathbf{p}_2 - \hat{e}\hat{k})D(\mathbf{r}_2, \mathbf{r}_1|\varepsilon)] \times[(2\mathbf{e}\cdot\mathbf{p}_1 + \hat{e}\hat{k})\tilde{D}(\mathbf{r}_1, \mathbf{r}_2|\varepsilon')]\}. \tag{40}$$

Here and below, we assume the subtraction from the integrand of its value at $Z\alpha = 0$. For calculations in the leading approximation in m/ε, the function $D(\mathbf{r}_2, \mathbf{r}_1|\varepsilon)$ can be used in the form [12]

$$D(\mathbf{r}_2, \mathbf{r}_1|\varepsilon) = \left[1 + \frac{\boldsymbol{\alpha}\cdot(\mathbf{p}_1 + \mathbf{p}_2)}{2\varepsilon}\right]D^{(0)}(\mathbf{r}_2, \mathbf{r}_1|\varepsilon), \tag{41}$$

where $D^{(0)}(\mathbf{r}_2, \mathbf{r}_1|\varepsilon)$ is the semiclassical Green function of the Klein–Gordon equation in the external field. The function $\tilde{D}$ is obtained from Eq. (41) by the replacement $D^{(0)} \longrightarrow D^{(0)*}$. Representation (41) can be used directly for the calculating the screening correction to the spectrum. It is shown below that it can be used for calculating the correction in m/ε as well.

Substituting Eq. (41) in Eq. (40) and taking the trace, we obtain

$$\frac{d\sigma^{\gamma}}{d\omega} = \frac{2\alpha\omega}{\varepsilon^2}\mathrm{Re}\int\!\!\int d\mathbf{r}_1 d\mathbf{r}_2 \exp(-i\mathbf{k}\cdot\mathbf{r}) \times \sum_{\lambda_\gamma}\Big\{4[\mathbf{e}\cdot\mathbf{p}_2 D_2^{(0)}][\mathbf{e}\cdot\mathbf{p}_1 D_1^{(0)}] + \frac{\omega^2}{\varepsilon\varepsilon'}[\mathbf{e}\cdot(\mathbf{p}_1+\mathbf{p}_2)D_2^{(0)}][\mathbf{e}\cdot(\mathbf{p}_1+\mathbf{p}_2)D_1^{(0)}]\Big\}, \quad (42)$$

$$D_2^{(0)} = D^{(0)}(\mathbf{r}_2, \mathbf{r}_1|\varepsilon), \quad D_1^{(0)} = D^{(0)*}(\mathbf{r}_1, \mathbf{r}_2|\varepsilon').$$

In deriving Eq. (42), we integrated the terms containing second derivatives of $D^{(0)}$ by parts. We are interested in the Coulomb corrections that can be obtained from Eq. (42) by the additional subtraction of the Born term ($\propto(Z\alpha)^2$) from the integrand.

3.1. Next-to-Leading Term in m/ε for Coulomb Corrections to the Spectrum

We start with Eq. (40) and introduce the variables

$$\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1, \quad \boldsymbol{\rho} = \frac{\mathbf{r}\times[\mathbf{r}_1\times\mathbf{r}_2]}{r^2}, \quad z = -\frac{\mathbf{r}\cdot\mathbf{r}_1}{r^2}. \quad (43)$$

We note that the variable $\boldsymbol{\rho}$ in this section has a quite different meaning than the variable $\boldsymbol{\rho}$ in the representation for $\mathbf{A}(\boldsymbol{\Delta})$ in the previous section (see Eq. (11)). The analysis performed shows that the leading contribution to the term under discussion originates from the region $\rho \sim 1/m$ and $\theta, \psi \sim m/\varepsilon \ll 1$, where θ is the angle between the vectors $\mathbf{r}_2$ and $-\mathbf{r}_1$, and ψ is the angle between the vectors $\mathbf{r}$ and $\mathbf{k}$. Screening can then be neglected, and we can use the semiclassical Green function D in a Coulomb field obtained in [9],

$$D(\mathbf{r}_2, \mathbf{r}_1|\varepsilon) = \frac{i\kappa e^{ikr}}{8\pi^2 r_1 r_2}\int d\mathbf{q}\exp\Big[i\frac{\kappa r q^2}{2r_1 r_2}\Big]\Big(\frac{2\sqrt{r_1 r_2}}{|\mathbf{q}-\boldsymbol{\rho}|}\Big)^{2iZ\alpha\lambda} \times\Big\{\Big(1+\frac{\lambda r}{2r_1 r_2}\boldsymbol{\alpha}\cdot\mathbf{q}\Big)\Big(1+i\frac{\pi(Z\alpha)^2}{2\kappa|\mathbf{q}-\boldsymbol{\rho}|}\Big) - \frac{\pi(Z\alpha)^2}{4\kappa^2}(\gamma^0\lambda - \boldsymbol{\gamma}\cdot\mathbf{r}/r)\frac{\boldsymbol{\gamma}\cdot(\mathbf{q}-\boldsymbol{\rho})}{|\mathbf{q}-\boldsymbol{\rho}|^3}\Big\}, \quad (44)$$

$$\lambda = \mathrm{sgn}\,\varepsilon, \quad \kappa = \sqrt{\varepsilon^2 - m^2}, \quad \boldsymbol{\alpha} = \gamma^0\boldsymbol{\gamma},$$

where $\mathbf{q}$ is a two-dimensional vector in the plane perpendicular to $\mathbf{r}$. We note that, because the angle θ is small, we can assume that the variable z belongs to the interval (0, 1) and $r_1 = rz$, $r_2 = r(1-z)$. The function $\tilde{D}$ entering Eq. (40) is obtained from Eq. (44) by the replacement $\kappa \longrightarrow -\kappa$ and $\lambda \longrightarrow -\lambda$. The contribution of the last term in braces in Eq. (44) vanishes after taking the trace in Eq. (40). Therefore, this term can be omitted in the problem under consideration. The remaining terms in Eq. (44) can be represented in form (41) with

$$D^{(0)}(\mathbf{r}_2, \mathbf{r}_1|\varepsilon) = \frac{i\kappa e^{i\kappa r}}{8\pi^2 r_1 r_2}\int d\mathbf{q}\exp\Big[i\frac{\kappa r q^2}{2r_1 r_2}\Big] \times\Big(\frac{2\sqrt{r_1 r_2}}{|\mathbf{q}-\boldsymbol{\rho}|}\Big)^{2iZ\alpha\lambda}\Big(1+i\frac{\pi(Z\alpha)^2}{2\kappa|\mathbf{q}-\boldsymbol{\rho}|}\Big). \quad (45)$$

Then, using the relation

$$(\mathbf{e}\cdot\mathbf{p}_{1,2})D^{(0)}(\mathbf{r}_2, \mathbf{r}_1|\varepsilon) = \frac{i\kappa^2 e^{i\kappa r}}{8\pi^2 r_1 r_2} \times\int d\mathbf{q}\exp\Big[i\frac{\kappa r q^2}{2r_1 r_2}\Big]\Big(\frac{2\sqrt{r_1 r_2}}{|\mathbf{q}-\boldsymbol{\rho}|}\Big)^{2iZ\alpha\lambda} \times\Big(1+i\frac{\pi(Z\alpha)^2}{2\kappa|\mathbf{q}-\boldsymbol{\rho}|}\Big)\Big(\mp\frac{\mathbf{e}\cdot\mathbf{r}}{r}+\frac{\mathbf{e}\cdot\mathbf{q}}{r_{1,2}}\Big), \quad (46)$$

and passing from the variables $\mathbf{r}_{1,2}$ to the variables $\mathbf{r}$, $\boldsymbol{\rho}$, and z, we obtain from (42) that

$$\frac{d\sigma_C^{\gamma}}{d\omega} = -\frac{\alpha\omega\varepsilon'}{32\pi^4\varepsilon}\mathrm{Re}\int\frac{d\mathbf{r}}{r^5}\int_0^1\frac{dz}{z^2(1-z)^2}\int\!\!\int\!\!\int d\mathbf{q}_1 d\mathbf{q}_2 d\boldsymbol{\rho} \times\exp\Big[\frac{i\omega r}{2}\Big(\psi^2+\frac{m^2}{\varepsilon\varepsilon'}\Big)+i\frac{\varepsilon q_1^2-\varepsilon' q_2^2}{2rz(1-z)}\Big] \times\Big\{\Big(\frac{Q_2}{Q_1}\Big)^{2iZ\alpha}-1+2(Z\alpha)^2\ln^2\frac{Q_2}{Q_1} + \frac{i\pi(Z\alpha)^2}{2}\Big[\Big(\frac{Q_2}{Q_1}\Big)^{2iZ\alpha}-1\Big]\Big(\frac{1}{\varepsilon Q_1}-\frac{1}{\varepsilon' Q_2}\Big)\Big\} \times\sum_{\lambda_\gamma}\Big\{4\varepsilon\varepsilon'\Big(-\mathbf{e}\cdot\mathbf{r}+\frac{\mathbf{e}\cdot\mathbf{q}_1}{1-z}\Big)\Big(\mathbf{e}\cdot\mathbf{r}+\frac{\mathbf{e}\cdot\mathbf{q}_2}{z}\Big) + \frac{\omega^2}{z^2(1-z)^2}(\mathbf{e}\cdot\mathbf{q}_1)(\mathbf{e}\cdot\mathbf{q}_2)\Big\}, \quad (47)$$

where $Q_{1,2} = |\mathbf{q}_{1,2} - \boldsymbol{\rho}|$. The integral over $\boldsymbol{\rho}$ can be taken

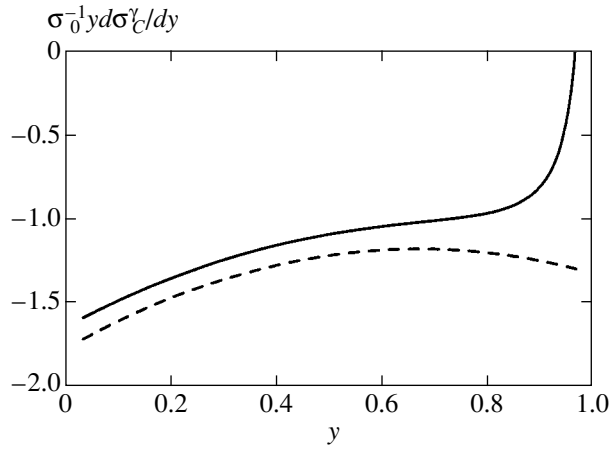


Fig. 3. The dependence of $\sigma_0^{-1} y d\sigma_C^\gamma/dy$ on y (see (49)) for $Z = 82$, $\varepsilon = 50$ MeV. Dashed curve: leading approximation; solid curve: first correction is taken into account.

using the relations (see Appendix B in [9])

$$f(Z\alpha) = \frac{1}{2\pi(Z\alpha)^2 q^2} \times \int d\boldsymbol{\rho}\left[\left(\frac{Q_2}{Q_1}\right)^{2iZ\alpha} - 1 + 2(Z\alpha)^2 \ln^2\frac{Q_2}{Q_1}\right] = \mathrm{Re}[\psi(1+iZ\alpha) + C], \quad (48)$$

$$g(Z\alpha) = \frac{i}{4\pi q}\int \frac{d\boldsymbol{\rho}}{Q_2}\left[\left(\frac{Q_2}{Q_1}\right)^{2iZ\alpha} - 1\right] = Z\alpha\frac{\Gamma(1-iZ\alpha)\Gamma(1/2+iZ\alpha)}{\Gamma(1+iZ\alpha)\Gamma(1/2-iZ\alpha)},$$

where

$$\psi(t) = d\ln\Gamma(t)/dt,$$

$C = 0.577\ldots$ is the Euler constant, and $q = |\mathbf{q}_1 - \mathbf{q}_2|$. We next perform summation over the photon polarization, pass to the variables

$$\tilde{\mathbf{q}} = \mathbf{q}_1 + \mathbf{q}_2, \quad \mathbf{q} = \mathbf{q}_1 - \mathbf{q}_2,$$

and take all integrals in the following order: $d\Omega_{\mathbf{r}}$, $d\tilde{\mathbf{q}}$, $d\mathbf{q}$, dr, dz. The final result for Coulomb corrections to the bremsstrahlung spectrum is given by

$$y\frac{d\sigma_C^\gamma}{dy} = -4\sigma_0\left[\left(y^2 + \frac{4}{3}(1-y)\right)f(Z\alpha) - \frac{\pi^3(2-y)m}{8(1-y)\varepsilon}\left(y^2 + \frac{3}{2}(1-y)\right)\mathrm{Re}\,g(Z\alpha)\right], \quad (49)$$

$$y = \omega/\varepsilon, \quad \sigma_0 = \alpha(Z\alpha)^2/m^2.$$

In this formula, the term proportional to $f(Z\alpha)$ corresponds to the leading approximation [3], and the term proportional to $\mathrm{Re}\,g(Z\alpha)$ is an $O(m/\varepsilon)$-correction. In our recent paper [9], this result was obtained by means of the substitution rules from the spectrum of pair production by photon in a Coulomb field. Formula (49) describes bremsstrahlung from electrons. For the spectrum of photons emitted by positrons, it is necessary to change the sign of $Z\alpha$ in (49). The $O(m/\varepsilon)$-correction becomes especially important in the hard part of the spectrum, as can be seen in Fig. 3, where $\sigma_0^{-1} y d\sigma_C^\gamma/dy$ with the correction (solid line) and without it (dashed line) are shown for $Z = 82$ and $\varepsilon = 50$ MeV. We note that, in the whole range of y, the relative magnitude of the correction is appreciably larger than m/ε due to the presence of a large numerical coefficient.

3.2. Screening Corrections

In this subsection, we calculate the screening correction to the high-energy asymptotics of $d\sigma_C^\gamma/d\omega$, considering λ_C/r_{scr} as a small parameter.

We start from Eq. (42) and use the semiclassical Green function $D^{(0)}(\mathbf{r}_2, \mathbf{r}_1|\varepsilon)$ for an arbitrary localized potential $V(\mathbf{r})$. This Green function was obtained in [10] with the first correction in m/ε taken into account. The leading term has the form (see also [12])

$$D^{(0)}(\mathbf{r}_2, \mathbf{r}_1|\varepsilon) = \frac{i\kappa e^{i\kappa r}}{8\pi^2 r_1 r_2} \times \int d\mathbf{q}\exp\left[i\frac{\kappa r q^2}{2r_2 r_2} - i\lambda r\int_0^1 dx V(\mathbf{r}_1 + x\mathbf{r} - \mathbf{q})\right]. \quad (50)$$

Similarly to Eq. (47), we obtain

$$\frac{d\sigma_C^\gamma}{d\omega} = -\frac{\alpha\omega\varepsilon'}{32\pi^4\varepsilon}\mathrm{Re}\int\frac{d\mathbf{r}}{r^5}\int_0^1\frac{dz}{z^2(1-z)^2}\iiint d\mathbf{q}_1 d\mathbf{q}_2 d\boldsymbol{\rho} \times \exp\left[i\Phi + \frac{i\omega r}{2}\left(\psi^2 + \frac{m^2}{\varepsilon\varepsilon'}\right) + i\frac{\varepsilon q_1^2 - \varepsilon' q_2^2}{2rz(1-z)}\right] \times \sum_{\lambda_\gamma}\left\{4\varepsilon\varepsilon'\left(-\mathbf{e}\cdot\mathbf{r} + \frac{\mathbf{e}\cdot\mathbf{q}_1}{1-z}\right)\left(\mathbf{e}\cdot\mathbf{r} + \frac{\mathbf{e}\cdot\mathbf{q}_2}{z}\right) + \frac{\omega^2}{z^2(1-z)^2}(\mathbf{e}\cdot\mathbf{q}_1)(\mathbf{e}\cdot\mathbf{q}_2)\right\}, \quad (51)$$

where

$$\Phi = r\int_0^1 dx[V(\mathbf{r}_1 + x\mathbf{r} - \mathbf{q}_2) - V(\mathbf{r}_1 + x\mathbf{r} - \mathbf{q}_1)]. \quad (52)$$

As we see in what follows, it is meaningful to retain the screening correction only in the case where $r_{scr} \ll \Delta_{\min}^{-1}$, which is considered below. Then, the main contribution to integral (51) comes from the region

$$1/m \lesssim \rho \lesssim r_{scr} \ll r$$

and

$$q_{1,2} \sim 1/m.$$

Under these conditions, the narrow region

$$\delta x = \rho/r \ll 1$$

around the point

$$x_0 = -\frac{\mathbf{r}_1 \cdot \mathbf{r}}{r^2} = z$$

is important in the integration over x in Eq. (52). Therefore, we can perform this integration from $-\infty$ to ∞. The phase Φ then becomes

$$\Phi = 2Z\alpha\ln(Q_2/Q_1) + \Phi^{(scr)} = 2Z\alpha\ln(Q_2/Q_1) + r\int_{-\infty}^{\infty} dx[\delta V(\mathbf{r}_1 + x\mathbf{r} - \mathbf{q}_2) - \delta V(\mathbf{r}_1 + x\mathbf{r} - \mathbf{q}_1)], \quad (53)$$

where $\delta V(\mathbf{r})$ is the difference between the atomic potential and the Coulomb potential of a nucleus. The notation in Eqs. (51) and (53) is the same as in Eq. (47). It is seen that

$$\Phi_{scr} \sim \rho\delta V(\rho) \sim \frac{Z\alpha\delta V(\rho)}{V(\rho)} \ll 1 \quad \text{for} \quad \rho \sim m$$

and

$$\Phi_{scr} \sim \frac{q_{1,2}}{\rho} \sim \frac{1}{m\rho} \ll 1 \quad \text{for} \quad \rho \sim r_{scr} \gg \frac{1}{m}.$$

Therefore, expression (51) can be expanded in $\Phi^{(scr)}$. In our calculation of the screening correction $d\sigma_C^{\gamma(scr)}/d\omega$, we retain the linear term of the expansion in $\Phi^{(scr)}$. The function $\delta V(\mathbf{R})$ can be expressed via the atomic electron form factor $F(\mathbf{Q})$ as

$$\delta V(\mathbf{R}) = \int\frac{d\mathbf{Q}}{(2\pi)^3}\exp(i\mathbf{Q}\cdot\mathbf{R})F(\mathbf{Q})\frac{4\pi Z\alpha}{Q^2}. \quad (54)$$

Substituting this formula in Eq. (53) and taking the integral over x from $-\infty$ to ∞, we obtain

$$\Phi^{(scr)} = \int\frac{d\mathbf{Q}_\perp}{(2\pi)^2}[\exp(i\mathbf{Q}_\perp\cdot(\boldsymbol{\rho}-\mathbf{q}_2)) - \exp(i\mathbf{Q}_\perp\cdot(\boldsymbol{\rho}-\mathbf{q}_1))]F(\mathbf{Q}_\perp)\frac{4\pi Z\alpha}{Q_\perp^2}, \quad (55)$$

where $\mathbf{Q}_\perp$ is a two-dimensional vector lying in the plane perpendicular to $\mathbf{r}$. We next use the identity (see Eqs. (22) and (23) in [13])

$$\int d\boldsymbol{\rho}\left(\frac{|\boldsymbol{\rho}-\mathbf{q}_2|}{|\boldsymbol{\rho}-\mathbf{q}_1|}\right)^{2iZ\alpha}\exp[i\mathbf{Q}_\perp\cdot(\boldsymbol{\rho}-\mathbf{q}_{1,2})] = \frac{q^2}{4Q_\perp^2}\int d\mathbf{f}\left(\frac{f_2}{f_1}\right)^{2iZ\alpha}\exp[i\mathbf{q}\cdot\mathbf{f}_{1,2}/2], \quad (56)$$

where

$$\mathbf{q} = \mathbf{q}_1 - \mathbf{q}_2, \quad \mathbf{f}_{1,2} = \mathbf{f} \mp \mathbf{Q}_\perp.$$

Expanding the exponential in Eq. (51) with respect to $\Phi^{(scr)}$ and using relation (56), we take the integrals over $\mathbf{q}_{1,2}$, $\mathbf{r}$, and z and obtain

$$y\frac{d\sigma_C^{\gamma(scr)}}{dy} = \frac{4\alpha(Z\alpha)}{\pi}\mathrm{Im}\int\frac{d\mathbf{Q}_\perp}{Q_\perp^4}F(\mathbf{Q}_\perp)\int\frac{d\mathbf{f}}{2\pi} \times\left[\left(\frac{f_2}{f_1}\right)^{2iZ\alpha} - 2iZ\alpha\ln\frac{f_2}{f_1}\right]\left[\frac{S(\xi_1)}{f_1^2} - \frac{S(\xi_2)}{f_2^2}\right],$$

$$S(\mu) = \frac{(\mu-1)}{\mu^2}\left\{\frac{1}{2\sqrt{\mu}}[y^2(3-\mu)] + (y-1)(\mu^2+2\mu-3)\ln\left[\frac{\sqrt{\mu}+1}{\sqrt{\mu}-1}\right] - 3y^2 - (y-1)(\mu-3)\right\}, \quad (57)$$

$$y = \frac{\omega}{\varepsilon}, \quad \xi_{1,2} = 1 + \frac{16m^2}{f_{1,2}^2}.$$

Using the trick introduced in [13], we can rewrite this formula in another form. We multiply the integrand in (57) by

$$1 \equiv \int_{-1}^{1} dx\,\delta\left(x - \frac{2\mathbf{f}\cdot\mathbf{Q}_\perp}{\mathbf{f}^2+\mathbf{Q}_\perp^2}\right) = (\mathbf{f}^2+\mathbf{Q}_\perp^2) \times\int_{-1}^{1}\frac{dx}{|x|}\delta((\mathbf{f}-\mathbf{Q}_\perp/x)^2 - \mathbf{Q}_\perp^2(1/x^2-1)), \quad (58)$$

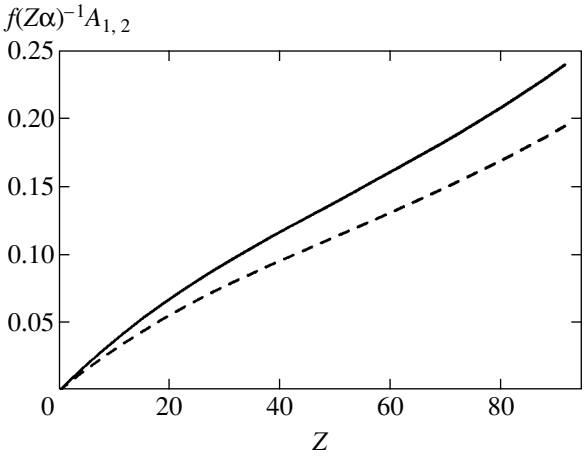


Fig. 4. The dependence of $A_1/f(Z\alpha)$ (solid curve) and $A_2/f(Z\alpha)$ (dashed curve) on Z.

change the order of integrations over $\mathbf{f}$ and x, and make the shift

$$\mathbf{f} \longrightarrow \mathbf{f} + \mathbf{Q}_\perp/x.$$

After that, the integration over f can be easily performed. Then, we make the substitution $x = \tanh\tau$ and obtain

$$y\frac{d\sigma_C^{(scr)}}{dy} = 16\sigma_0 m^2 \int_0^\infty \frac{d\mathbf{Q}_\perp}{2\pi}\frac{F(\mathbf{Q}_\perp)}{Q_\perp^4} \times \int_0^\infty \frac{d\tau}{\sinh\tau}\left[\frac{\sin(2Z\alpha\tau)}{2Z\alpha} - \tau\right]\int_0^{2\pi}\frac{d\varphi}{2\pi}[e^\tau S(\mu_2) - e^{-\tau}S(\mu_1)], \quad (59)$$

$$\mu_{1,2} = 1 + \frac{8m^2 e^{\mp\tau}\sinh^2\tau}{Q_\perp^2(\cosh\tau + \cos\varphi)}.$$

According to Eq. (57), the correction $yd\sigma_C^{\gamma(scr)}/dy$ has the form

$$y\frac{d\sigma_C^{\gamma(scr)}}{dy} = \sigma_0[A_1(1-y) + A_2y^2]. \quad (60)$$

Shown in Fig. 4 is the Z-dependence of the ratio $A_{1,2}/f(Z\alpha)$ calculated numerically with the use of form factors from [14]. For the less realistic Yukawa potential, we can perform analytic calculations of the functions A_i. It turns out that their dependence on the parameter $\beta = \lambda_c/r_{scr}$ has the form

$$A_i = (Z\alpha)^2\beta^2(a_i\ln^2\beta + b_i\ln\beta + c_i), \quad (61)$$

where b_i and c_i are some functions of $Z\alpha$, while a_i does not depend on $Z\alpha$. Recalling that β is proportional to $Z^{1/3}$ in the Thomas–Fermi model, we see that A_i depend on Z mainly via the factor

$$(Z\alpha)^2\beta^2 \propto (Z\alpha)^2 Z^{2/3}.$$

Therefore, it is quite natural that $yd\sigma_C^{\gamma(scr)}/dy$ calculated with the use of the exact form factors is well fitted by the expression

$$y\frac{d\sigma_C^{\gamma(scr)}}{dy} \approx 8.6\times 10^{-3}\sigma_0(Z\alpha)^2 \times Z^{2/3}[1.2(1-y) + y^2]. \quad (62)$$

In fact, the accuracy of this fit for all Z is better than a few percent.

It follows from Eq. (61) that, for $r_{scr} \gtrsim \Delta_{\min}^{-1}$, the factor β^2 in the screening correction is extremely small, $\beta^2 \lesssim (m/\varepsilon)^2$. The terms of such an order were systematically neglected in our consideration. Hence, within our accuracy, the account of the screening correction is meaningful only for $r_{scr} \ll \Delta_{\min}^{-1}$.

4. CONCLUSIONS

We have performed a detailed analysis of Coulomb corrections both to the differential and the integrated cross sections of bremsstrahlung in an atomic field. We have calculated the next-to-leading term in the high-energy asymptotics of the bremsstrahlung spectrum. Similar to the leading term of the high-energy asymptotics of Coulomb corrections to the spectrum, this term is independent of screening in the leading order in the parameter λ_c/r_{scr}. We have also calculated the first correction to the spectrum in the parameter λ_c/r_{scr}.

We have shown that, in contrast to Coulomb corrections to the spectrum, Coulomb corrections to the differential cross section strongly depend on screening even in the leading approximation. This dependence is very important in the region that gives the main contribution to the integral over $\Delta_\perp$. We have performed the explicit integration over $\Delta_\perp$ of $d\sigma_C^\gamma$ for arbitrary screening and have verified the independence of the final result from screening.

We also examined the effect of the finite beam size on Coulomb corrections to bremsstrahlung in a Coulomb field of a heavy nucleus. Similar to the effect of screening, the finiteness of the beam size leads to a strong modification of Coulomb corrections to the differential probability, while the probability integrated over $\Delta_\perp$ depends only on the density of the electron beam at zero impact parameter.

ACKNOWLEDGMENTS

This work was supported in part by the Russian Foundation for Basic Research (project no. 03-02-16510) and the Russian Science Support Foundation.

REFERENCES

1. H. A. Bethe and W. Heitler, Proc. R. Soc. London, Ser. A **146**, 83 (1934).
2. V. B. Berestetskii, E. M. Lifshitz, and L. P. Pitaevskii, *Quantum Electrodynamics*, 2nd ed. (Nauka, Moscow, 1980; Pergamon Press, Oxford, 1982).
3. H. A. Bethe and L. C. Maximon, Phys. Rev. **93**, 768 (1954).
4. H. A. Olsen, L. Maximon, and H. Wergeland, Phys. Rev. **106**, 27 (1957).
5. H. A. Olsen, Phys. Rev. **99**, 1335 (1955).
6. H. A. Olsen and L. Maximon, Phys. Rev. **114**, 887 (1959).
7. H. A. Olsen, Phys. Rev. D **68**, 033008 (2003).
8. R. N. Lee, A. I. Milstein, and V. M. Strakhovenko, hep-ph/0307388.
9. R. N. Lee, A. I. Milstein, and V. M. Strakhovenko, Phys. Rev. A **69**, 022708 (2004).
10. R. N. Lee, A. I. Milstein, and V. M. Strakhovenko, Zh. Éksp. Teor. Fiz. **117**, 75 (2000) [JETP **90**, 66 (2000)].
11. R. N. Lee and A. I. Milstein, Phys. Rev. A **61**, 032103 (2000).
12. R. N. Lee and A. I. Milstein, Phys. Lett. A **198**, 217 (1995).
13. R. N. Lee, A. I. Milstein, and V. M. Strakhovenko, Phys. Rev. A **57**, 2325 (1998).
14. J. H. Hubbell and I. Øverbø, J. Phys. Chem. Ref. Data **8**, 69 (1979).

Journal of Experimental and Theoretical Physics, Vol. 100, No. 1, 2005, pp. 106–115.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 1, 2005, pp. 120–130.

ELECTRONIC PROPERTIES OF SOLIDS

Fragmentation of Bands and Fermi Surfaces in Cuprate Stripe Phases

M. Ya. Ovchinnikova

Semenov Institute of Chemical Physics, Russian Academy of Sciences, ul. Kosygina 4, Moscow, 119991 Russia
e-mail: movchin@center.chph.ras.ru

Received July 13, 2004

Abstract—The band structure and evolution of the Fermi surfaces of stripe phases were studied using the *t*–*t*'–*U* Hubbard model in the mean field approximation. The appearance of quasi-one-dimensional "impurity" subbands caused by the localization of particles on domain walls inside the Hubbard gap is confirmed. Among vertical stripe phases parallel to *y* bonds, the Y_8 and Y_4 structures with distances l = 8a and 4a between domain walls were found to be stable. Fermi surface segments in antinodal or nodal directions were shown to correspond to an "impurity" band or the main band related to the entire antiferromagnetic domain region. This is a probable explanation of the difference in the properties of ARPES spectra at different Fermi surface regions observed for $La_{2-x}Sr_xCuO_4$. It was shown for the Y_8 structure that the topology of the Fermi surface changed and an isotropic pseudogap opened at the point corresponding to a p = 1/8 doping level. Attempts at relating this property to the anomalous suppression of T_c in LSCO at p = 1/8 encountered difficulties. The low dispersion of the impurity band and the wide gap separating it from the lower Hubbard band in diagonal stripe phases formed at $p < 0.05$ create prerequisites for the existence of the insulating state at nonzero doping.

1. INTRODUCTION

Angle-resolved photoemission spectroscopy (ARPES) [1, 2] directly probes quasi-particles in cuprates. For instance, the superconducting gap was measured and the existence of a pseudogap revealed by this method in the normal state of underdoped cuprates based on Bi (BSCCO). Like the superconducting gap of *d* symmetry, the pseudogap is absent in the nodal diagonal directions (on the line of the superconducting gap nodes) and opens in the antinodal quasi-momentum directions close to the points $M(\pm\pi, 0)$, $(0, \pm\pi)$. There are many arguments [1–3] in favor of the conclusion that the pseudogap is determined by the properties of the antiferromagnetically correlated Fermi liquid, namely, by the energy profile of the lower Hubbard band. This profile is well discernible in the ARPES data on insulating undoped antiferromagnets. At a low doping, it determines hole pockets in the nodal directions around the point $(\pi/2, \pi/2)$ and the pseudogap in the excitation spectrum in the antinodal quasi-momentum directions. The homogeneous solutions of the *t*–*t*'–*U* or *t*–*t*'–*J* models substantiate this picture at low *h*-doping and predict a change of the topology of the Fermi surface from small-sized hole pockets to a large electronic-type Fermi surface as the doping increases.

The observed evolution of the Fermi surface as the degree of *h* doping increases, however, contradicts the assumption of the formation of a *e*–type Fermi surface. For $La_{2-x}Sr_xCuO_4$ (LSCO), the Fermi surface arc remains hole-like at substantial doping levels and always intersects the $M(0, \pi)$–$Y(\pi, \pi)$ segment. It follows from the ARPES data on LSCO [4–6] that its Fermi surface has segments of two different types. The first type is straight Fermi surface segments parallel to CuO bonds close to the *M* point. The second type is segments close to the nodal diagonal directions with systematic photoemission intensity suppression on them. This picture was obtained for underdoped LSCO by photoemission with photons of energy $h\nu$ = 29 eV [4–6] in the main Brillouin zone. Measurements at $h\nu$ = 55 eV with a higher resolution in the second Brillouin zone showed that, along with bright Fermi surface segments in the antinodal directions, there were Fermi surface segments along the boundary of the magnetic Brillouin zone with a narrow momentum distribution and a low integral intensity [7, 8]. This behavior was related in [7] to the formation of stripe structures, that is, antiphase antiferromagnetic domains arranged in line. The suggestion was made that the transition between two types of structures, from diagonal stripe structures to stripe structures parallel to CuO bonds, occurred in the region of the transition from the insulating to superconducting state of LSCO at low doping levels ($\delta \sim 0.03$–0.07).

The observation of incommensurate peaks in inelastic neutron scattering at $Q = (\pi, \pi) \pm (\delta, \pm\delta)$ or $Q = (\pi \pm \delta, \pi)$, $(\pi, \pi \pm \delta)$ in the insulating or superconducting phase, respectively (see [9–11]), lends support to the existence of such quasi-static structures. A comparative study of the incommensurate peaks from superlattices in elastic neutron scattering and incommensurate peaks

in inelastic neutron scattering favors the suggestion that the latter originate from slowly fluctuating stripe structures [12]. The pinning of such structures caused by doping with zinc [13] or under magnetic field action [14] substantiates this conclusion. Additional broadening of quasi-particles at the points $k = (k_0, \pi - k_0)$ ($k_0 \sim 0.175\pi$), where the Fermi surface turned from the direction $M(\pi, 0)$–$M'(0, \pi)$ toward segments parallel to $\Gamma(0, 0)$–$M(\pi, 0)$ [7], was also related to stripe phases.

The Fermi surface in electronically doped cuprates is substantially different from that in *h*-doped samples. The observed Fermi surface in the underdoped region bounds electron pockets around the points $M(\pm\pi, 0)$, $(0, \pm\pi)$ [15] in consistency with the energy profile of the upper Hubbard band [16, 17]. Further doping generates a fragmentary Fermi surface with an additional segment in the diagonal direction [18]. Such a patched character of the Fermi surface was explained in [19] by the formation of electron and hole pockets simultaneously from the upper and lower Hubbard subbands of the *t*–*t*'–*U* model. In [20], the hypothesis according to which the fragmentary structure of the Fermi surface was related to stripe phases was also discussed.

The purpose of this work is to show that the fragmentary Fermi surface structure and several other special features of bands and spectral intensities could appear as a result of the formation of stripe phases. We show that Fermi surface fragmentation for a periodic stripe structure is caused by splitting the initial band into several subbands, as distinct from splitting into two (lower and upper) Hubbard subbands in the homogeneous antiferromagnetic state. This is accompanied by the splitting off of "impurity" quasi-one-dimensional subbands present inside the Hubbard gap of the homogeneous antiferromagnetic state. These in-gap bands are related to particles localized on domain walls. We show that different Fermi surface segments correspond either to the particles that belong to the impurity band and are associated with domain walls or to particles from the entire region of antiferromagnetic domains.

Our analysis is based on self-consistent solutions to the Hubbard model obtained by the mean field method. The results are in agreement with those reported in [20, 21] and based on a simplified description of stripe phases, which were modeled by a handpicked potential. The mean field method allows the stability of stripe structures to be estimated in a crude approximation, and unrealistic hypothetical structures, for instance, stripe structures with charge modulation and long-range antiferromagnetic spin order simultaneously, to be excluded from consideration. Our calculations substantiate many of the properties of quasi-particle excitations in stripe phases observed in model calculations and calculations by the mean field method [20–24]. In particular, they lend support to the interpretation of quasi-one-dimensional quasi-particle states inside the Hubbard gap as states localized on domain walls.

We consider the *t*–*t*'–*U* Hubbard model with band energy ϵ_k and one-center interaction $U > 0$ ($t = 1$ and all energies are in *t* units throughout). The model does not include long-range Coulomb interactions but gives a quite correct description of the properties of quasi-particles characterized by large quasi-momenta near the Fermi boundary. The Hamiltonian of the model in the mean field approximation is insufficient for describing the superconducting state and must be augmented by empirical interaction of electrons on neighboring centers. The corresponding equations should be derived with more accurate inclusion of correlations. We, however, use the original *t*–*t*'–*U* Hubbard model to study stripe structures in the normal state. The procedure for seeking a solution by the mean field method with a periodic field structure is quite standard [24]. The periodic structure is determined by the two-dimensional translation vectors E_i and B_i in the direct and reciprocal lattices, respectively,

$$\mathbf{E}_i = (E_{ix}, E_{iy}), \quad \mathbf{B}_i = (B_{ix}, B_{iy}), \quad i = 1, 2, \quad (1)$$

which satisfy the conditions $E_iB_j = 2\pi\delta_{ij}$.

An arbitrary site $n = n(L, j) = E_1L_1 + E_2L_2 + j$ of the periodic structure with n_c centers per unit cell is described by the coordinates $L = (L_1, L_2)$ and $j = (j_x, j_y)$ of the unit cell and the site inside it, respectively. A quasi-momentum $k \in G$ inside the Brillouin zone of the original lattice is expressed via the momentum $\tilde{k}$ reduced to the Brillouin zone $\tilde{G}$ of the periodic structure, $k = \tilde{k} + B_1m_1 + B_2m_2$. The set of integers m_1, m_2 numbers all independent Umklapp vectors. The $\tilde{G}$ and *G* regions are bounded by the conditions $|\tilde{k}\,B_i| \le |B_i|^2/2$ and $|k_{x(y)}| \le \pi$, respectively.

The role of the order parameters for the periodic mean field solutions is played by electron densities and mean spin vectors of each unit cell site,

$$r_j = \langle r_{n(L,j)} \rangle, \quad S_{\alpha j} = \langle S_{\alpha, n(L,j)} \rangle. \quad (2)$$

Here, averaging is over all $N_L = N/n_c$ unit cells and α numbers spin vector components.

The mean energy is found in the mean field approximation as an explicit function of order parameters (2) [24], and the wave function is determined by the population of the one-electron eigenstates $\chi^\dagger_{k\lambda}$ of the linearized Hamiltonian

$$H_{\rm lin} = T + N_L\sum_j \{2Ur_j\hat{r}_j - 2US_{\mu j}\hat{S}_{\mu j}\} = \sum_{\tilde{k} \in \tilde{G}} \tilde{h}_{\tilde{k}}.$$

This Hamiltonian is partitioned into independent contributions for each reduced quasi-momentum $\tilde{k}$ value. Here, *T* is the kinetic energy and $\hat{r}_j$ and $\hat{S}_{\mu j}$ are the operators corresponding to one-electron means (2). The

eigenstates of the Hamiltonian H_{lin} are expanded in the momentum representation into a set of $2n_c$ Fermi operators

$$\chi^{\dagger}_{\tilde{k}\lambda} = \sum_{m,\sigma} c^{\dagger}_{\tilde{k}+Bm,\sigma} W_{m\sigma,\lambda}(\tilde{k}), \tag{3}$$

where $\lambda = 1, \ldots, 2n_c$, $Bm = B_1m_1 + B_2m_2$, and the vectors $\tilde{k} + Bm$ span the entire phase space G.

The matrix of eigenvectors $W_{m\sigma,\lambda}$ and eigenvalues $E_{\tilde{k},\lambda}$ are found by diagonalizing the $h_{\tilde{k}}$ matrix in the basis set $\{c^{\dagger}_{\tilde{k}+Bm,\sigma}\}$,

$$(h_{\tilde{k}})_{m,\sigma,m',\sigma'} W_{m',\sigma',\lambda} = W_{m,\sigma,\lambda} E_{\tilde{k},\lambda}. \tag{4}$$

Here,

$$\begin{aligned}(h_{\tilde{k}})_{m,\sigma,m',\sigma'} &= \delta_{mm'}\delta_{\sigma\sigma'}\epsilon_{\tilde{k}+Bm} \\ &+ U\sum_j \varphi(j, m'-m)[r_j\delta_{\sigma\sigma'} - S_{\mu j}(\sigma_\mu)_{\sigma\sigma'}],\end{aligned} \tag{5}$$

where $\varphi(j, m) = \exp[iBmj]$ and $j = (j_x, j_y)$ runs over all n_c centers of the unit cell. Order parameters (2) themselves are in turn calculated from the matrix of eigenvectors W and Fermi functions f as

$$\begin{aligned}\{r_j, S_{\alpha j}\} &= \frac{1}{2N}\sum_{\tilde{k}\in\tilde{G}}\sum_{m,s,m',s'}\{\sigma_0, \sigma_\alpha\}_{ss'}\varphi(j, m'-m) \\ &\times W^{*}_{m,s,\lambda}(\tilde{k}) W_{m',s',\lambda}(\tilde{k}) f(E_{\tilde{k}\lambda} - \mu).\end{aligned} \tag{6}$$

The Pauli matrices σ_α and σ_0 in (6) correspond to the $S_{\alpha j}$ and r_j components, respectively. Equations (4) and (6) determine the self-consistent solutions of the mean field method for a periodic field structure.

The intensity of the photoemission of electrons with the momentum projection k onto the ab plane and energy $E = h\nu - \omega$ is

$$I(k, \omega) = |M(k)|^2 A(k\omega) f(\omega) \otimes R_{\omega k}. \tag{7}$$

It is determined by the matrix element $M(k)$, spectral density $A(k\omega)$, and Fermi function f and is accompanied by the convolution with the Gaussian function $R_{\omega k}$ [25] that imitates finite resolution in k and energy. The dependence of the matrix element M on k was studied in [26]. Here, we use a constant value of M, because our goal is to study the influence of the structure on the spectral density A.

In the one-electron approximation,

$$\begin{aligned}A(k, \omega) &= \frac{1}{N} \\ &\times \sum_{\tilde{k}\in\tilde{G}}\sum_{m,\sigma,\lambda} \left|W_{m,\sigma,\lambda}(\tilde{k})\right|^2 \bar{\delta}(E_{\tilde{k}\lambda} - \mu - \omega)\delta_{k,\tilde{k}+Bm}.\end{aligned} \tag{8}$$

Here, $\lambda = 1, \ldots, 2n_c$ and $m = (m_1, m_2)$ number independent Umklapp vectors $Bm = B_1m_2 + B_2m_2$. The standard replacement of the δ function in (8) by a function with a finite width Ω is implied. The map $I(k_x, k_y, \omega = 0)$ visualizes both the main and shadow Fermi surface segments. Although band energies are periodic functions in the k space, intensity (7) and spectral density (8) are devoid of such periodicity. For this reason, even if the matrix element in (7) is independent of k, various Fermi surface regions manifest themselves with different intensities because of the compound nature of band operators in the presence of a structure.

Apart from the maps $I(k_x, k_y, \omega = 0)$, we consider intensity maps averaged over a certain frequency window $2\Delta\omega$,

$$I_{\Delta\omega}(k, \omega) = \int d\omega' I(k\omega') R\left(\frac{\omega - \omega'}{\Delta\omega}\right). \tag{9}$$

Here, R is the corresponding Gaussian function with width $\Delta\omega$ that imitates finite resolution in ω. The construction of such maps requires function normalization by its maximum value. It follows that the brightness and width of Fermi boundaries on them depends on the width of the frequency window $\Delta\omega$ in (9). In particular, at a large width $\Delta\omega$, not only the true Fermi boundaries with sharp population steps but also boundaries with smoother $n(k)$ variations are seen in the map of the $I_{\Delta\omega}$ function. Such regions correspond to the opening of a insulating gap or pseudogap.

Lastly, if $k = k(l)$ is varied along a certain contour, the map of the $A(k, \omega)$ function on the plane of two variables $k(l)$ and ω reproduces the dispersion of the energy levels $E_\lambda(k)$ along the selected contour. Each level is then weighted with a weight factor $|W_{k,\sigma,\lambda}|^2$ corresponding to the contribution $c^{\dagger}_{k\sigma}$ to band eigenstate (3). To see all the $E_\lambda(k)$ bands as a whole rather than their "bright" fragments only, it is sufficient to construct the map of function (8) with replacing the $|W|^2$ factor by one.

2. THE STRUCTURE OF VERTICAL STRIPE DOMAINS

Among the stripe phases with domains aligned with the y-oriented CuO bonds, the structures with domain walls centered on x bonds, wall spacings $l = 4$, 6, 8, and 10, and $n_c = 2l$ (l is in lattice parameters) were calculated. Let these phases be named $Y_4, \ldots, Y_{10}$, respectively. Without adjustment of model parameters, the mean field approximation ignoring interaction with the lattice describes a tendency in structure changes rather than the exact order of their sequence as the doping varies.

The mean energies $\bar{H}(p)$ per lattice site are shown in Fig. 1 as functions of the doping $p = 1 - n$ for the

hole-doped model with the parameters $U = 4$ and $t' = 0.1$ (in t units) for a series of structures and for the homogeneous antiferromagnetic state. The latter has the lowest energy when $p \longrightarrow 0$. In agreement with the earlier results [24], stripe states are stabler in hole-doped models than the homogeneous antiferromagnetic state. The Y_{10}, Y_8, Y_6, and Y_4 structures sequentially become stablest as p increases. Starting with certain $p > p^*(l)$ inside the interval l, two additional domain walls appear in the Y_8 and Y_{10} structures, and the width of the true antiferromagnetic domain decreases to $4a$ and $5a$. The p^* values correspond to kinks in the dependence of the mean energy on the doping and lie in the region where Y_4 is the stablest structure. At large p, the Y_4 structure with $\Delta l = 4$ is obviously distinguished as the stablest structure.

Figure 1 also displays variations in the electron and spin densities $r(n)$ and $d(n) = (-1)^{n_x + n_y}\langle S_z(n)\rangle$ on sites along the n_x coordinate parallel to CuO bonds for some structures in the hole-doped system with $U = 4$ and $t' = 0.1$. These variations show the degree of charge localization and spin density distortion close to domain walls. The self-consistent periodic spin field $F(n) = U(-1)^n d(n)$ (or the spin structure factor) is somewhat different from the handpicked potential with rectangular modulation [20, 21]. At low t', it actually contains only two harmonics F_q with the vectors $q = Bm$ closest to the vector $Q = (\pi, \pi)$ of the homogeneous antiferromagnetic state.

The main and shadow Fermi surface segments on the complete phase space plane $|k_{x(y)}| \le \pi$ for the Y_4 stripe structure with $l = 4a$ and $n_c = 8$ aligned with y bonds are seen in Fig. 2a, where the map of the intensity of photoemission $I(k, \omega = 0)$ at the doping level $p = 0.15$ is shown. The Fermi surfaces of the stripe structure are substantially different from those of the homogeneous antiferromagnetic solution for the model with the same doping level. The Fermi surfaces for the latter are boundaries along hole pockets prolate along the boundary of the magnetic Brillouin zone. The periodic field of spins of antiphase antiferromagnetic domains forms quasi-one-dimensional Fermi surface segments perpendicular to the direction of stripe domains. The map of intensities $I(k_x, k_y, \omega = 0)$ symmetrized along the x and y stripe domain orientations is shown in Fig. 2b. The absence of a Fermi boundary in the diagonal direction means a pseudogap opening in this direction instead of the pseudogap in the region of points $M = (0, \pi), (\pi, 0)$ for a homogeneous solution.

The intensity $I_{\Delta\omega}$ map averaged over the frequency window $\Delta\omega = 0.08t$ according to Eq. (9) is shown in Fig. 2c. In addition to the bright Fermi surface segments, the broadening of $\Delta\omega$ makes visible quasi-particles with energies exceeding the pseudogap width in the diagonal direction.

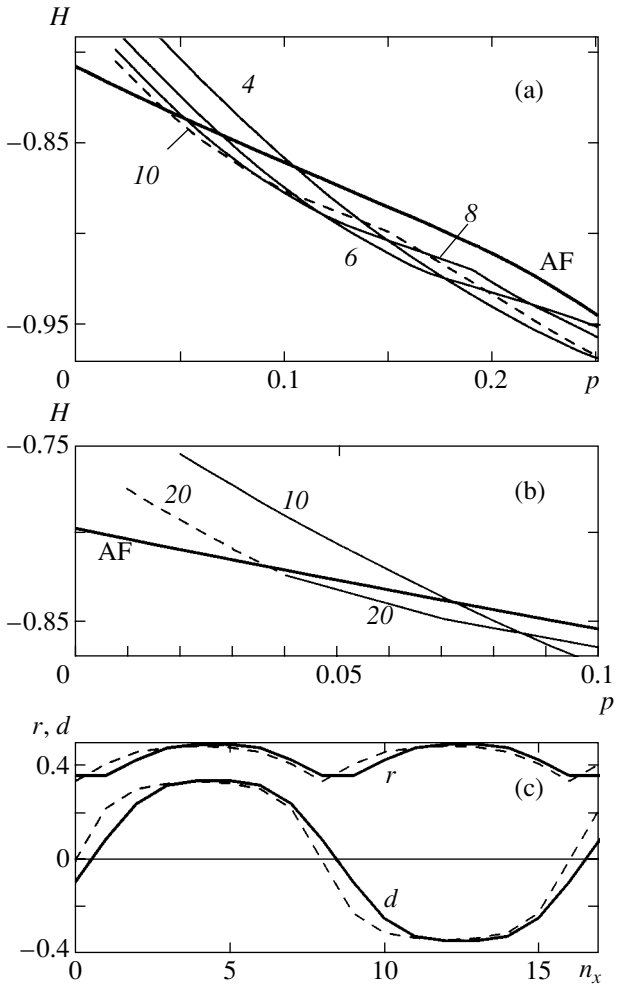


Fig. 1. (a) Doping $p = 1 - n$ dependences of the mean energy per node for the hole-doped model with the parameters $U = 4.0$ and $t' = 0.1$ of vertical stripe structures and homogeneous antiferromagnetic state (thick curve AF). The curves for Y_l structures are labeled by the corresponding $l = 4, 6, 8$, and 10 values. (b) The same at a low p value for the diagonal stripe phases D_{10} and D_{20}. The dashed lines correspond to approximate mean field results in the region of incomplete convergence. (c) Mean density and spin (2) variations depending on the coordinate n_x of lattice sites for the Y_8 (solid lines) and D_8 (dashed lines) structures.

The band structure of periodic stripes and spectral weight transfer between subbands are seen in the spectral density $A(k, \omega)$ maps on the (k, ω) plane as k changes along a certain contour. The band energies weighted according to (8) and the system of the same band levels $E_\lambda(k) - \mu$ of the periodic structure with unit weights are shown in Fig. 3 as functions of k, which varies along the path $\Gamma(0, 0)$–$M(\pi, 0)$–$Y(\pi, \pi)$. The $E_\lambda(k)$ energies [but not the intensities $A(k)$] are, as they should be, periodic functions of k with the period $2\pi/l$ in the contour regions normal to the direction of stripe domains. In reality, only a small number of nonshadow band levels contribute to the spectral function for each

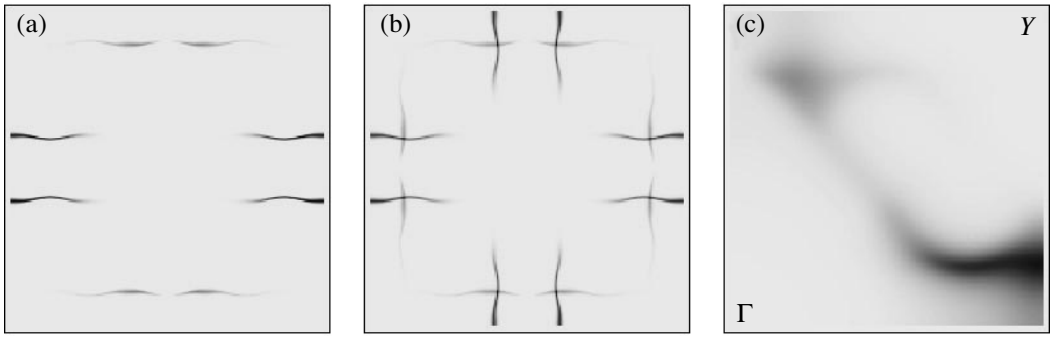


Fig. 2. (a) Intensity map [Eq. (7)] on the $|k_{x,y}| \le \pi$ plane representing the Fermi surface of the Y_4 structure of vertical stripes with $l = 4$ for the model with $U = 4$ and $t' = 0.1$; doping level $p = 0.15$. (b) The same map averaged over the x- and y-oriented stripe structures. (c) Map of intensity (8) averaged over the frequency window $\Delta\omega = 0.08t$ in the first phase space quadrant $0 < k_{x(y)} \le \pi$.

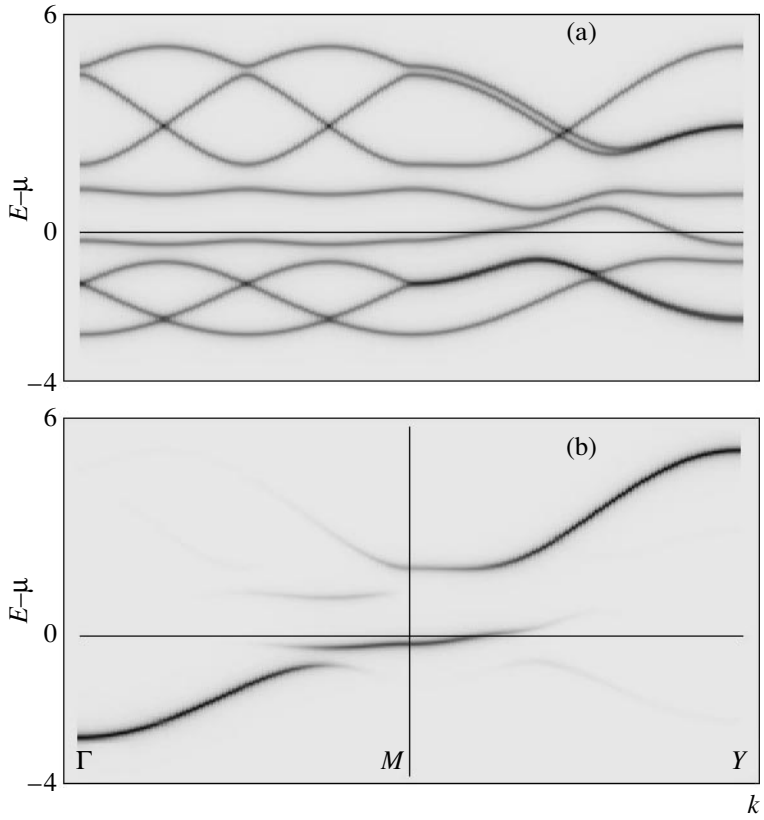


Fig. 3. (a) Maps of spectral density (8) revealing band energies $E_\lambda(k) - \mu$ of the Y_4 stripe structure as k varies along the contour Γ–$M(\pi, 0)$–$Y(\pi, \pi)$. (b) The same bands from the density map of levels with unit weights in (8). Model parameters and doping are as in Fig. 2.

k and, accordingly, manifest themselves in photoemission. In agreement with the results obtained in [20, 21, 24], it follows from Fig. 2 that the transfer of spectral weight between the subbands of periodic structure occurs in such a way that bright band regions tend to reproduce the unperturbed band. Bright band fragments close to the boundary of the magnetic Brillouin zone appear inside the Hubbard gap of the excitation spectrum of the homogeneous antiferromagnetic state.

Details of the behavior of "weighted" bands $E(k) - \mu$ in the vicinity of the boundary of the magnetic Brillouin zone as k varies along three sections, namely, Γ–$M(\pi, 0) \longrightarrow Y(\pi, \pi)$, Γ–$S(\pi/4, 3\pi/4) \longrightarrow Y$, and Γ–$(\pi/2, \pi/2) \longrightarrow Y$, are shown in Fig. 4. Along the last two sections, none of the bands intersects the chemical potential level, which means a pseudogap opening in these directions and is in agreement with the absence of a diagonal Fermi surface segment in Fig. 2b. The Fermi

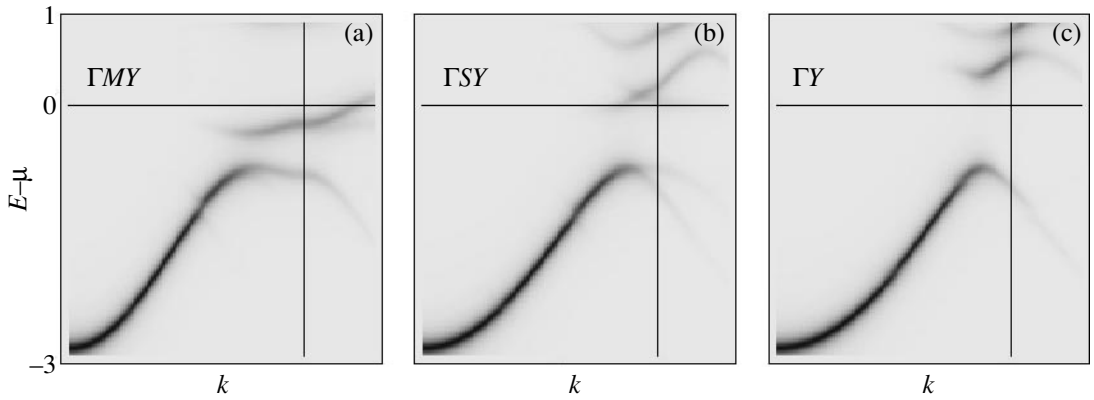


Fig. 4. Behavior of bands and spectral density transfer between them close to the boundary of the magnetic Brillouin zone from the $A(k, \omega)$ map; k varies along three contours Γ–$M(\pi, 0)$–$Y(\pi, \pi)$, Γ–$S(3\pi/4, \pi/4)$–Y, and Γ–Y in the region limited by the condition $k_x + k_y \le 1.3\pi$. The vertical lines correspond to the k values at the intersections of the contours with the magnetic Brillouin zone. Maps *b* and *c* display pseudogap opening at the corresponding points of this zone. The Y_4 structure, doping, and model are as in Figs. 2 and 3.

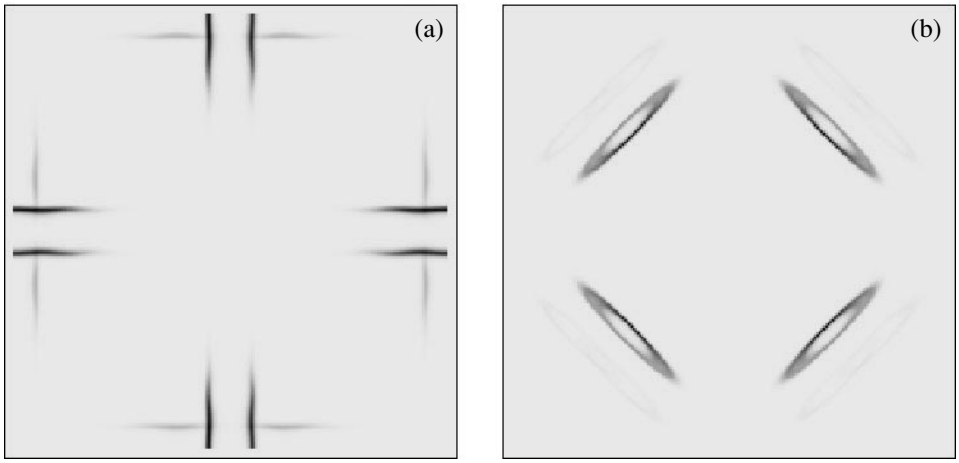


Fig. 5. Photoemission intensity $I(k, \omega = 0)$ maps revealing the Fermi surface for the Y_8 stripe structure (with averaging over the x and y orientations); maps (a) and (b) correspond to doping levels $p = 0.1 < 1/8$ and $p = 0.15 > 1/8$, respectively. Anomalous doping $p = 1/8$ corresponds to pseudogap opening along the boundary of the magnetic Brillouin zone.

surface also disappears near $M_y(0, \pi)$. According to the interpretation suggested in [20–22], there appears an impurity 1D band of quasi-particles localized on domain walls and moving along them. For vertical stripe domains, the dispersion of such a band is large, of order t.

Calculations of the Y_8 and Y_{10} structures show that the shape of the Fermi surface similar to that described above (Figs. 2a, 2b) only exists at the doping levels $p < 1/8$ and $p < 1/10$, respectively. At the same time, Fermi surface segments in the nodal diagonal directions appear at $p > 1/l$, and a pseudogap habitual with BSCCO again opens near the $M(\pi, 0)$, $(0, \pi)$ points. The Fermi surfaces of the Y_8 structure with the doping levels $p = 0.1 < 1/8$ and $p = 0.15 > 1/8$ are shown in Fig. 5. At the very instant of the change of the topology of the Fermi surface, at the doping level $p = 1/8$, a pseudogap opens along the entire boundary of the magnetic Brillouin zone. The impurity band lies wholly above the chemical potential level μ and is fully free, whereas the remaining seven lower levels of the main band originating from states delocalized over the antiferromagnetic domain lie above μ at all k values and are fully populated. The gap between them at the chemical potential level is seen in Fig. 6, which displays the spectral density $A(k, E - \mu)$ maps as k varies along three sections, namely, $\Gamma - M(\pi, 0)$–$Y(\pi, \pi)$, Γ–$S(\pi/4\ 3\pi/4)$–Y, and Γ–Y, and the corresponding unweighted eigenvalues. The arrangement of the latter explains the scheme of filling the main and impurity in-gap bands. The appearance of a diagonal Fermi surface segment at $p > 1/8$ in the Y_8 structure means the formation of a hole pocket in the remaining main band, which contains seven levels

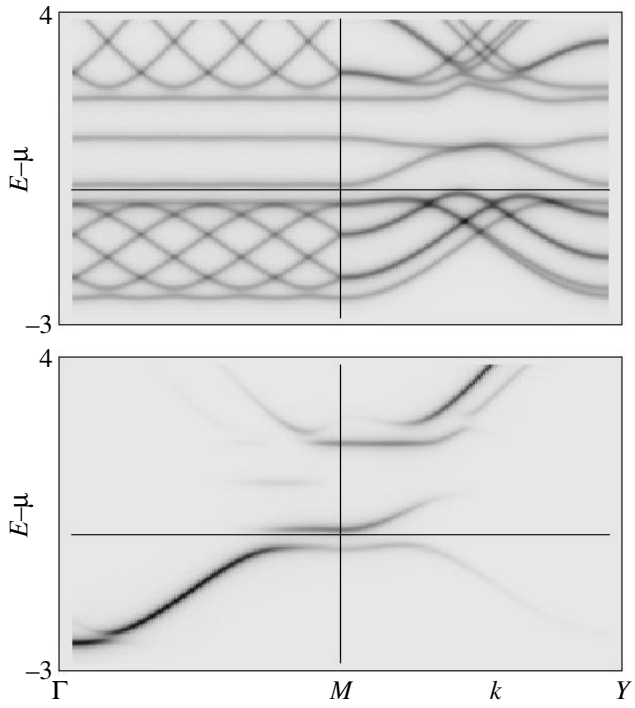


Fig. 6. Behavior of bands $E_\lambda(k) - \mu$ for the Y_8 structure at anomalous doping $p = 1/8$ on the maps of the spectral (bottom) and unweighted (top) densities of states. Quasi-momentum k changes along the Γ–$M(\pi, 0)$–Y contour. The horizontal line is the chemical potential level.

per domain layer, while the impurity band that split off from the main band inside the Hubbard gap is completely depopulated. Solutions for the Y_{10} structure exhibit similar change of the Fermi surface topology at $p = 1/10$. At this p value, the mean energy of the Y_{10} structure is, however, higher than that of the Y_8 structure.

It is more realistic to describe LSCO as an ensemble of several quasi-static structures with domains of various widths or mean charge density fluctuations in space or time. Simplified model calculations of a disordered stripe phase were performed in [20]. Taking into account spatial charge fluctuations or (and) structure variations, we can expect quasi-one-dimensional Fermi surface segments in antinodal directions and Fermi surface segments in nodal directions to manifest themselves simultaneously in the ARPES data. It follows from our analysis that different Fermi surface segments have different origins, and antinodal segments originate from impurity band states localized on domain walls, whereas Fermi surface segments in nodal directions originate from states delocalized over the antiferromagnetic domain. Clearly, the former are more sensitive to the influence of defects, impurities, and structure variations. This could explain the difference in the shape of ARPES signals from LSCO for nodal and antinodal Fermi surface segments (the presence versus absence of sharp peaks in the energy distribution of photoelectrons [6, 7] for the corresponding segments).

It also follows from the band fragmentation described above that the pseudogap region of the magnetic Brillouin zone with a suppressed photoemission intensity is retained in the region of the intersection of the standard Fermi surface arc and the boundary of the magnetic Brillouin zone in the vicinity of the point $S(\pi/4, 3\pi/4)$. Such magnetic Brillouin zone points are called hot points in [7], because, starting with these points, the observed broadening of the photoemission peak is interpreted as the switching on of an additional mechanism of scattering of particles with the corresponding k value. According to the concept of stripe phases, the disappearance of the Fermi surface in this direction is caused by special features of the bands, namely, the retention of the pseudogap in the region between two Fermi surface segments of different natures. Near hot points, spectral weight is transferred from band states delocalized over the antiferromagnetic domain to an impurity quasi-one-dimensional band, which is more sensitive to the stripe structure period. As has been mentioned, peak broadening in the energy distribution of photoelectrons in antinodal directions starting with hot points could be a consequence of a disordered arrangement of domain walls.

An attempt can be made to relate the anomaly of the electronic properties of the Y_8 structure at $p = 1/8$ to the anomalous suppression of superconductivity and T_c in LSCO at this p value. However, a contradiction remains between the positions of inelastic neutron scattering peaks $q = \pi(1, 1 \pm 1/8)$ expected for the stripe structure Y_8 and their positions $q = \pi(1, 1 \pm 1/4)$ observed for LSCO at $p = 1/8$. The latter favor the Y_4 structure.

Studies of other periodic structures showed that, in the mean field approximation, charge modulation (charge density waves) is only possible in the presence of antiphase antiferromagnetic domains with boundaries between them. The artificial order parameters that describe charge density waves for in-phase antiferromagnetic domains do not survive: the solution by the mean field method converges to a homogeneous antiferromagnetic solution. This is also true of the periodic variant of bubble-type structures [20] with a 4×4 cell and two-dimensional structures with antiphase antiferromagnetic domains of size 4×4 and charge modulation along the x and y axes. A consideration of the latter in the mean field approximation leads to a homogeneous paramagnetic solution. At certain parameter values, there exist dimeric structure states with spin currents on plaquettes 4 and mean spin vectors lying in the ab plane (32 centers per unit cell). Their energy is higher than that of the vertical stripe structure Y_4, and their Fermi surface in no way corresponds to the segments of the observed Fermi surface.

3. DIAGONAL STRIPE PHASES

The differences between diagonal stripe structures and the structures aligned with bonds, which were considered above, are of interest. The diagonal structures with a distance of $l/\sqrt{2}$ between domain walls and the number $n_c = 2l$ sites per unit cell will be denoted by D_l. The presence of such structures in the insulating LSCO phase follows from inelastic neutron scattering peaks with $Q = (\pi, \pi) \pm (\delta, \pm\delta)$ at a doping level of $p < 0.05$ [9–11]. We calculated D_l structures with the translation vectors $E_1 = (-1, 1)$ and $E_2 = (l, l)$ and domain walls centered on sites rather than bonds, because the energy of diagonal structures increases when domain walls are centered on bonds. The energies of the D_{10} and D_{20} structures with the numbers of sites $n_c = 2l = 20$, 40 per unit cell are shown in Fig. 1b relative to the energy of the antiferromagnetic state at $p < 0.1$. The data presented in Fig. 1 and the experimental results reported in [9–11] lead us to expect the existence of diagonal stripe phases D_l with $l > 20$ in the low doping region ($p \le 0.05$). Nevertheless, the characteristics of the Fermi surface and the impurity band inside the Hubbard gap will be illustrated as an example of the D_8 stripe structure with a smaller number of levels.

The Fermi surface map and the behavior of bands along the sections Γ–$Y(\pi, \pi)$ and Γ–$Y'(-\pi, \pi)$ for the diagonal D_8 stripe structure with the translation vectors $E_1 = (-1, 1)$ and $E_2 = (8, 8)$, $n_c = 16$, and domains parallel to E_1 are shown in Fig. 7. As was predicted in [20, 21], the quasi-one-dimensional Fermi surface segments are perpendicular to the direction of stripe domains. An impurity in-gap band with a small dispersion appears in the middle of the Hubbard gap. As follows from the reasoning of the authors of [20, 22], the dispersion of this band is proportional to the small hopping t', as distinct from a similar in-gap band in vertical stripes with width $\sim t$. At such a small dispersion of the in-gap impurity band and a wide gap separating it from the remaining lower Hubbard band, doping and stripe structure period variations enable a system having the insulating state even at nonzero doping $p_h \neq 0$ to be modeled [22].

The spectral density $A(k, \omega)$ map is shown in Fig. 8 along two diagonal sections for the periodic structure D_8 with $n_c = 16$, $E_2 = (8, 8)$, and a doping level of $p = 1/8$. It corresponds with the structure of "weighted" bands along two diagonals, parallel and perpendicular to stripe domains. The plots at the bottom display the system of "unweighted" one-electron levels. Along the entire boundary of the magnetic Brillouin zone, a wide gap separates the major part of the lower Hubbard band from the impurity band of states split off and localized along domain walls. At $p = 1/8$, the impurity band of the D_8 structure is empty. It follows that, at such a doping, the D_8 structure would have the properties of a insulating with a wide gap separating the conduction band. This situation is similar to that with the anomalous behavior of the vertical stripe structure Y_8 at doping $p = 1/8$. The difference is that the gap of excitations in the Y_8 structure would be narrow because of a large impurity band dispersion.

The question of whether or not the D_8 structure with $p = 1/8$ can exist, that is, the question of its stability as compared to vertical stripe structures, can hardly be answered unambiguously on the basis of mean field calculations. It follows from the positions of the low-frequency incommensurate peaks of inelastic neutron scattering by LSCO [9, 10] that we can expect the formation of diagonal stripe structures D_l with $l \ge 20$ only at a very low doping level, $p < 0.05$.

Any diagonal D_l structure at the corresponding doping level $p = 1/l$ retains the properties of the D_8 structure described above; that is, it remains insulating and has an equally wide gap separating the main filled band from the empty impurity conduction band. It follows that, at arbitrarily low doping in the region $p < 0.05$, where diagonal structures D_l with $l > 20$ are stabler, we

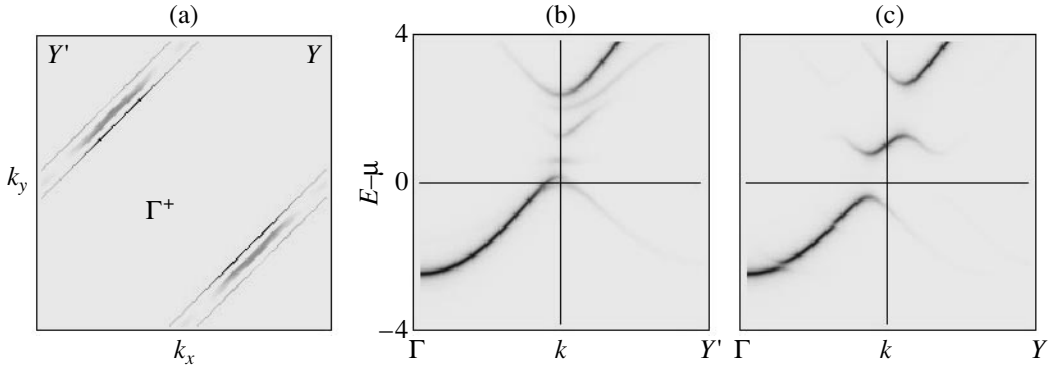


Fig. 7. (a) Intensity map for the D_8 structure with stripe domains parallel to the vector $B_2 = (-\pi, \pi)$ at the $p = 0.15$ doping level. The Fermi surface segments normal to the direction of domains are only seen. (b, c) Behavior of bands for k changing along the diagonals $\Gamma(0, 0)$–$Y'(-\pi, \pi)$ and Γ–$Y(\pi, \pi)$, respectively.

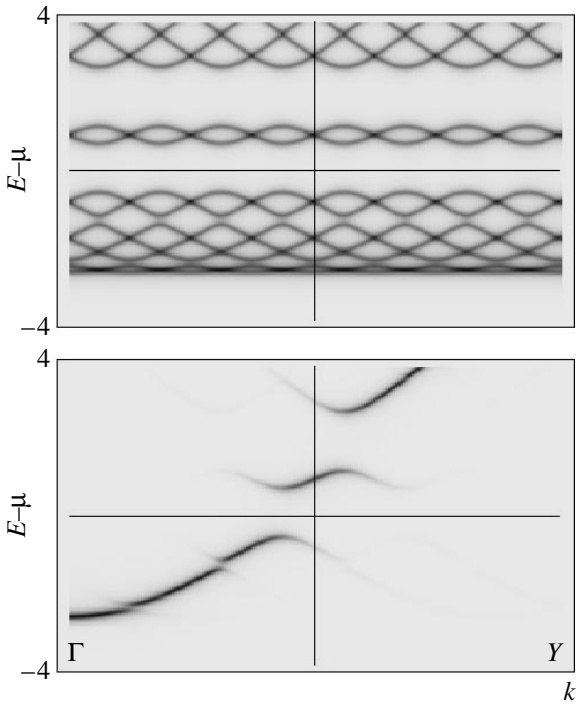


Fig. 8. Spectral density $A(k, \omega)$ maps (bottom) and band energies $E(k) - \mu$ (top) of the D_8 stripe structure at the $p = 1/8$ doping level for k changing along the Γ–Y diagonal normal to the direction of stripe domains.

always observe the formation of a set of diagonal domain walls with the mean distance $\bar{l} > 20/\sqrt{2}$ between them that gives rise to the insulating state of the system through splitting off the required number of empty impurity bands from the remaining fully filled main band. These arguments, first presented in [22], explain the existence of a insulating phase in LSCO at nonzero doping ($p < 0.05$).

As distinct from diagonal structures, vertical stripe domains are characterized by a large impurity band dispersion, $\Delta E \sim t$, on the order of the Hubbard gap width itself. Because of this and because of the dispersion of the edge of the lower Hubbard band, such structures cannot be responsible for the insulating state of the system over a substantial range of doping. The anomaly at $p = 1/8$ is an exception for the vertical stripe structure Y_8.

Lastly, note one more problem that arises in the description of the electron-doped $Nd_{2-x}Ce_xCuO_4$ cuprate (NCCO) at $p_e > 0.13$. This is the interpretation of the patched character of the Fermi surfaces observed in the maps of the intensity of the ARPES signal from NCCO [18]. The Fermi surfaces consist of the boundaries of the electron pockets around the $M(\pi, 0)$, $(0, \pi)$ points and additional Fermi surface segments in the diagonal nodal direction. This phenomenon was explained in [19] by the coexistence of electron pockets from the population of the upper Hubbard band and hole pockets in the neighborhood of $k \sim (\pi/2, \pi/2)$ from partial depopulation of the lower Hubbard band. Attempts at relating the appearance of the additional Fermi surface segments to stripe structures did not lead to an unambiguous result [20]. We also tried to find stripe structures that would reproduce such segments. In particular, the Fermi surface of the diagonal stripe structure D_{10} does have additional segments in the nodal direction perpendicular to the direction of diagonal stripe domains. The energy of this state for the *e*-doped Hubbard model is, however, higher than the energy of the homogeneous antiferromagnetic state, as with the *e*-doped models of the other stripe structures studied earlier [24].

To summarize, the formation of stripe structures and related Umklapp processes and spectral weight transfer between different subbands result in the fragmentation of bands and Fermi surfaces observed in photoemission. A quasi-one-dimensional impurity band is split off inside the Hubbard gap. It corresponds to states localized along domain walls that separate antiphase antiferromagnetic stripe domains. In the region of stability of stripe structures Y_l aligned with bonds and having domain widths $l = 4$, 6, 8, and 10, it is the impurity band that is responsible for the existence of Fermi surface segments aligned with bonds in antinodal directions at doping levels $p < 1/l$. At the same time, Fermi surface segments in nodal directions owe their existence to band states delocalized over antiferromagnetic domain sites and appear at doping levels $p > 1/l$. The different origins of the nodal and antinodal Fermi surface segments explains the difference in the characteristics of photoemission from these segments observed in the ARPES studies of LSCO [4–6]. For the Y_8 structure, the $p = 1/8$ critical value itself signifies the opening of a pseudogap along the entire magnetic Brillouin zone boundary. The suggestion that the Y_8 structure is responsible for anomalous suppression of T_c in LSCO at a $p = 1/8$ doping is, however, inconsistent with the size $l = 4$ of stripe domains that follows from neutron scattering measurements. Calculations of diagonal stripe structures substantiate the special features of impurity bands in this situation. The small dispersion of the impurity band and the wide gap separating it from the main band noticed in [20–22] are responsible for the existence of a insulating phase at nonzero doping. It would be useful to discuss the photoemission spectra of LSCO bearing in mind the details of the behavior of bands and Fermi surfaces of particular structures obtained in mean field calculations.

Note added in proof. Substantial progress in analysis of stripe phases was achieved by using the Gutzwiller approximation in [27, 28], where the dependences of the period, chemical potential, and transport properties of the structures on doping were explained.

ACKNOWLEDGMENTS

This work was financially supported by the Russian Foundation for Basic Research (project no. 03-03-32141). The author thanks V.Ya. Krivnov for valuable comments and help.

REFERENCES

1. A. Damascelli, Z.-X. Shen, and Z. Hussain, Rev. Mod. Phys. **75**, 473 (2003); cond-mat/0208504.
2. Z. M. Shen and D. S. Dessau, Phys. Rep. **253**, 1 (1995).
3. T. Timusk and B. Statt, Rep. Prog. Phys. **62**, 61 (1999).
4. T. Yoshida, X. J. Zhou, T. Sasagawa, *et al.*, Phys. Rev. Lett. **91**, 027001 (2003).
5. A. Ino, C. Kim, M. Nakamura, *et al.*, Phys. Rev. B **62**, 4137 (2000).
6. X. J. Zhou, T. Yoshida, S. A. Kellar, *et al.*, Phys. Rev. Lett. **86**, 5578 (2001).
7. X. J. Zhou, T. Yoshida, D. H. Lee, *et al.*, Phys. Rev. Lett. **92**, 187001 (2004).
8. X. J. Zhou, J. Shi, T. Yoshida, *et al.*, cond-mat/0405130.
9. S. A. Kivelson, I. P. Bindloss, E. Fradkin, *et al.*, Rev. Mod. Phys. **75**, 1201 (2003).
10. J. N. Tranquada, H. Woo, T. G. Perring, *et al.*, Nature **429**, 531 (2004).
11. M. Matsuda, M. Fujita, K. Yamada, *et al.*, Phys. Rev. B **62**, 9148 (2000).
12. M. Fujita, H. Goka, K. Yamada, *et al.*, cond-mat/0403396.
13. H. Kimura, M. Kofu, Y. Matsumoto, and K. Hirota, Phys. Rev. Lett. **91**, 067002 (2003).
14. B. Lake, H. M. Ronnow, N. B. Christensen, *et al.*, Nature **415**, 299 (2002).
15. N. P. Armitage, D. H. Lu, D. L. Feng, *et al.*, Phys. Rev. Lett. **86**, 1126 (2001).
16. A. A. Ovchinnikov, M. Ya. Ovchinnikova, and E. A. Plekhanov, Pis'ma Zh. Éksp. Teor. Fiz. **67**, 350 (1998) [JETP Lett. **67**, 369 (1998)]; Zh. Éksp. Teor. Fiz. **114**, 985 (1998) [JETP **87**, 534 (1998)]; **115**, 649 (1999) [**88**, 356 (1999)].
17. C. Kusko and R. S. Markiewicz, Phys. Rev. Lett. **84**, 963 (2000).
18. N. P. Armitage, D. H. Lu, C. Kim, *et al.*, Phys. Rev. Lett. **87**, 147003 (2001).
19. C. Kusko, R. S. Markiewicz, M. Lindroos, and A. Banzil, Phys. Rev. B **66**, 140513 (2002).
20. M. Granath, cond-mat/0401063.
21. M. Granath, V. Oganesyan, D. Orgad, and S. A. Kivelson, Phys. Rev. B **65**, 184501 (2002).
22. H. J. Shulz, J. Phys. (Paris) **50**, 2833 (1989).
23. V. J. Emery, S. A. Kivelson, and O. Zachar, Phys. Rev. B **56**, 6120 (1997).
24. A. A. Ovchinnikov and M. Ya. Ovchinnikova, Zh. Éksp. Teor. Fiz. **125**, 618 (2004) [JETP **98**, 546 (2004)]; cond-mat/0305685.
25. S. V. Borisenko, A. A. Kordyuk, S. Legner, *et al.*, Phys. Rev. B **64**, 094513 (2001).
26. M. Lindroos, S. Sahrakorpi, and A. Bansil, Phys. Rev. B **65**, 054514 (2002).
27. J. Lorenzana and G. Seibold, Phys. Rev. Lett. **89**, 136401 (2002); **90**, 066404 (2003).
28. G. Seibold and J. Lorenzana, Phys. Rev. B **69**, 134513 (2004).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 100, No. 1, 2005, pp. 116–125.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 1, 2005, pp. 131–140.

ELECTRONIC PROPERTIES OF SOLIDS

High-Frequency Response of Two-Well Nanostructures

V. F. Elesin

Moscow Engineering Physics Institute (State University), Moscow, 115409 Russia
e-mail: VEF@supercon.mephi.ru

Received September 1, 2004

Abstract—A consistent analytical theory is developed for coherent resonant electron tunnelling in a two-well nanostructure in the presence of a weak electromagnetic field. Simple expressions derived for the transmission coefficient and linear response of the two-well nanostructure make it possible to clarify the physics of processes and to express the gain as a function of the structure parameters. It is shown that the high-frequency response of the two-well nanostructure considerably exceeds the response of a one-well structure (resonance-tunnel diode) and that the application of a constant electric field makes it possible to tune the oscillation frequency and to increase the gain. It is concluded that two-well nanostructures can be used in designing terahertz oscillators. It is shown that, in contrast to a resonance-tunnel diode, interference of electrons between the wells and radiative "laser-type" transition play a decisive role in such structures. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Resonant tunnelling in nanostructures has wide prospects in designing integrated circuits, logic elements, oscillators, and lasers [1]. New striking possibilities appear when conditions for coherent tunnelling are created. Coherent lasers that do not require population inversion [2] and oscillators based on a two-band Stark ladder [3] may serve as examples of such devices.

At the present level of technology, coherence can be ensured over large lengths (up to a dozen quantum wells); this was demonstrated in experimental studies (see, for example, [4]). It should also be noted that the simple condition $\tau_\Gamma < \tau_{ph}$ for coherent tunnelling (τ_Γ is the residence time for an electron in a well and τ_{ph} is the time of coherence breakdown) may turn out to be too stringent. Indeed, the detailed theory shows that the electron–phonon interaction does not affect the attenuation of Bloch oscillations under certain conditions [5], while resonant tunnelling remains coherent even for $\tau_\Gamma \gg \tau_{ph}$ [6].

For this reason, it would be interesting to study the features of the interaction of an electromagnetic field with nanostructures under the conditions of coherent tunnelling. It should be noted that a theoretical description of coherent systems requires a consistent quantum-mechanical approach. This is so because resonant tunnelling is associated with quantum interference of electrons, which is exceptionally sensitive to the boundary conditions and the energy of the electrons being injected. Such an approach was used earlier for describing the high-frequency response of a single-well structure [7] (which is also referred to as a resonance-tunnel diode (RTD)).

This study aims at developing a theory of the high-frequency response for a two-well nanostructure (TWNS). Such a structure can be the simplest structure in a transition from an RTD to a superlattice with a Stark ladder. In addition, TWNS is of independent interest. First, it exhibits new features associated with interwell interference and level splitting. Second, the high-frequency response associated with transitions between split levels might be more intense than the response in an RTD [8, 9]. Thus, TWNSs are very promising for developing terahertz oscillators.

In this study, a consistent analytic theory of coherent resonant tunnelling of electrons interacting with an electromagnetic field is developed. A simple model of the structure and the approach used earlier in [7] are employed. Simple expressions derived here for the transmission coefficient and the response of the system make it possible to clarify the physics of the processes and to calculate the gain as a function of TWNS parameters.

In particular, it will be shown that the gain of a TWNS considerably exceeds the gain of an RTD, and that the application of a constant electric field makes it possible to retune the oscillation frequency. The conclusion that the gain increases in a certain interval of the constant field appears the most interesting. It should also be noted that in the same way as in [7] for an RTD, we managed to eliminate expressions of the type of $1/\omega^2$, which diverge for $\omega \longrightarrow 0$ and which are typical of approximate theories (see the literature cited in [7]), and obtain correct frequency dependences.

2. FORMULATION OF THE PROBLEM AND BASIC EQUATIONS

We will study a model of coherent tunneling in a TWNS using the approach developed in [7]. We consider a one-dimensional structure with δ-function barriers at points $x = 0$, a, $2a$ (see figure). A steady-state

electron flow proportional to q^2 is supplied to the first well from the left ($x \longrightarrow -\infty$); the energy ε of the electrons is approximately equal to the resonance level energy ε_R of the first well. In the region of the TWNS, a varying electric field $E(t)$ is applied and the field potential is

$$V(x, t) = 2U(x)\cos\omega t,$$
$$U(x) = \begin{cases} xU, & 0 < x < 2a, \\ 2Ua, & x > 2a, \quad U = -eE/2. \end{cases} \tag{1}$$

The resonance level of the second well may be displaced by V (for example, due to the action of the constant electric field). The wavefunction $\Psi(x, t)$ satisfies the Schrödinger equation

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

Here, we assume that $\hbar = 2m = 1$. We seek the steady-state solution to Eq. (2) in the form (see [7])

$$\Psi(x, t) = e^{-i\varepsilon t}[\Psi_{10}(x) + e^{-i\omega_n t}\Psi_{1n}(x)], \quad 0 \le x \le a,$$
$$\Psi(x, t) = e^{-i\varepsilon t}[\Psi_{20}(x) + e^{-i\omega_n t}\Psi_{2n}(x)], \quad a \le x \le 2a. \tag{3}$$

Partial wavefunctions Ψ_{l0}, Ψ_{ln}, $l = 1, 2$, $n = \pm 1$ describe electrons with quasienergies ε and $\varepsilon + \omega_n$, respectively.

Functions $\Psi_{l0}(x)$ in the zeroth approximation in the varying field satisfy the equations

$$p^2\Psi_{10}(x) + \Psi''_{10}(x) = 0, \quad p_1^2\Psi_{20} + \Psi''_{20} = 0 \tag{4}$$

and the boundary conditions

$$\Psi_{10}(0)(1-\beta) + \frac{\Psi'_{10}(0)}{ip} = 2q,$$
$$p^2 = \varepsilon; \quad \beta = \frac{\alpha}{ip},$$
$$\Psi_{20}(2a)(1-\beta_1) - \frac{\Psi'_{20}(2a)}{ip_1} = 0; \tag{5}$$
$$p_1^2 = p^2 + V; \quad \beta_1 = \frac{\alpha}{ip_1},$$
$$\Psi_{10}(a) = \Psi_{20}(a), \quad \Psi'_{20}(a) - \Psi'_{10}(a) = \alpha\Psi_{10}(a).$$

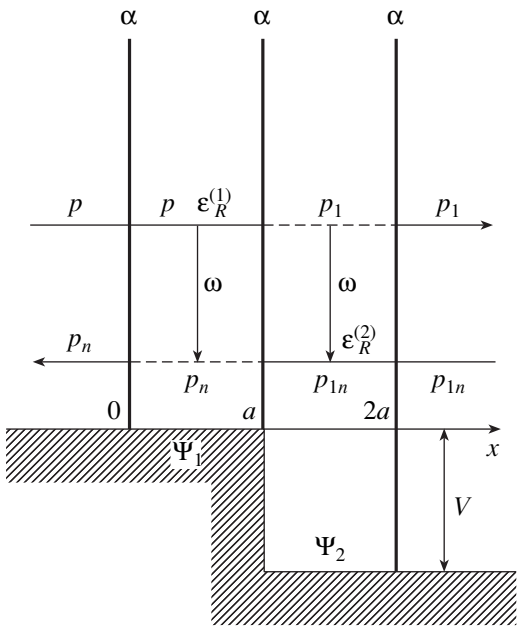


Figure.

The corresponding equations and boundary conditions for function $\Psi_{ln}(x)$ in the approximation linear in the field have the form

$$p_n^2\Psi_{1n} + \Psi''_{1n} = U(x)\Psi_{10}, \quad p_n^2 = p^2 - \omega_n, \quad \omega_n = -n\omega, \tag{6}$$
$$p_{1n}^2\Psi_{2n} + \Psi''_{2n} = U(x)\Psi_{20}, \quad p_{1n}^2 = p_1^2 - \omega_n;$$
$$\Psi_{1n}(0)(1-\beta_n) + \frac{\Psi'_{1n}(0)}{ip_n} = 0, \quad \beta_n = \frac{\alpha}{ip_n},$$
$$\Psi_{2n}(2a)(1-\beta_{1n}) - \frac{\Psi'_{2n}(2a)}{ip_{1n}} = 0, \quad \beta_{1n} = \frac{\alpha}{ip_{1n}}, \tag{7}$$
$$\Psi_{1n}(a) = \Psi_{2n}(a), \quad \Psi'_{2n}(a) - \Psi'_{1n}(a) = \alpha\Psi_{1n}(a).$$

The varying field induces the polarization currents (response)

$$J_l^c(x, t) = J_l^c(x)\cos\omega t,$$
$$J_l^s(x, t) = J_l^s(x)\sin\omega t.$$

Here, J^c is the current synphase with the field and J^s is the reactive current:

$$J_l^c(x) = \sum_{n=\pm 1} J_{l,n}^c, \tag{8}$$
$$J_{ln}^c = -ie[(\Psi_{l0}^*\Psi'_{ln} + \Psi_{ln}^*\Psi'_{l0}) - \text{c.c.}].$$

In the subsequent analysis, we will consider only current J^c leading to the emission (absorption) of the varying electric field.

3. WAVEFUNCTIONS AND TRANSMISSION COEFFICIENTS OF A TWO-WELL NANOSTRUCTURE IN ZERO VARYING FIELD

The solutions to Eqs. (4) have the form

$$\begin{aligned}\Psi_{10} &= A_{10}e^{ipx} + B_{10}e^{-ipx},\\ \Psi_{20} &= A_{20}e^{ip_1x} + B_{20}e^{-ip_1x},\end{aligned} \tag{9}$$

and the coefficients are given by

$$\begin{aligned}A_{10} &= \frac{2qe^{2ip_1a}\Delta_{10}(1)}{\Delta_{10}(2)}, \quad B_{10} = \frac{2qe^{2ipa}\Pi_{10}}{\Delta_{10}(2)},\\ A_{20} &= \frac{4q(2-\beta_1)e^{i(p-p_1)a}}{\Delta_{10}(2)},\\ B_{20} &= \frac{4q\beta_1 e^{i(3p_1+p)a}}{\Delta_{10}(2)},\end{aligned} \tag{10}$$

where the following notation is used:

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

$$\begin{aligned}\Pi_{10} &= (2-\beta_1)\left(\beta+1-\frac{p_1}{p}\right)\\ &\quad +\beta_1\left(\beta+\frac{p_1}{p}+1\right)e^{2ip_1a},\end{aligned} \tag{11}$$

$$\Delta_{10}(2) = (2-\beta)e^{2ip_1a}\Delta_{10}(1) - \beta e^{2ipa}\Pi_{10}.$$

Determinant $\Delta_{10}(2)$ describes the resonance properties of the TWNS and the form of the coefficients of transmission of electrons through the structure,

$$T(\varepsilon) = \frac{64}{|\Delta_{10}(2)|^2}. \tag{12}$$

Here and below, we consider the structure with "intense" barriers, for which $\alpha/p \gg 1$. It is in this limit that remarkable properties of quantum wells are fully manifested.

First, we will study the resonance properties of a symmetric TWNS, for which $V = 0$ and $p = p_1$. In this case, the determinant (11) assumes a simpler form $\Delta_{10}(2) \equiv \Delta_0(2)$,

$$\begin{aligned}\Delta_0(2) &= (2-\beta)^3 - 2\beta^2(2-\beta)e^{2ipa}\\ &\quad -\beta^2(2+\beta)e^{4ipa}.\end{aligned} \tag{13}$$

In the vicinity of the resonant values of energy, where $\Delta_0(2)$ assumes its minimal value, we can represent momentum p in the form

$$p = p_0 + \tilde{p}, \quad p_0 = \pi/a, \quad \tilde{p} \ll p_0.$$

In this case, we obtain the following equation for $x = \tilde{p}a$:

$$\begin{aligned}x^2(1+2ix) + 4\delta x\left(1+\frac{x}{\pi}\right) + 8ix^2\delta\left(1+\frac{x}{\pi}\right)\\ +3\delta^2\left(1+\frac{2x}{\pi}\right) - 2i\delta^3 = 0,\\ \delta = p_0/\alpha \ll 1.\end{aligned} \tag{14}$$

We will seek the solution to this equation in the form of expansion of $x = x_1 + x_2$ into a power series in small parameter δ. In the first approximation, we obtain the equation

$$x_1^2 + 4\delta x_1 + 3\delta^2 = 0, \tag{15}$$

which has the following solutions:

$$x_1^{(1)} = -\delta, \quad x_1^{(2)} = -3\delta. \tag{16}$$

Two solutions correspond to the splitting of the resonance level of an isolated well into two levels with energies

$$\varepsilon_R^{(1)} = p_0^2 - \frac{2p_0\delta}{a}, \quad \varepsilon_R^{(2)} = p_0^2 - \frac{6p_0\delta}{a} \tag{17}$$

and the splitting energy width

$$t_0 \equiv \varepsilon_R^{(1)} - \varepsilon_R^{(2)} = \frac{4p_0}{a}\delta = \frac{4p_0^2}{\alpha a}. \tag{18}$$

In the next approximation in δ, we have

$$x_2^{(1)} = \delta^2/\pi - i\delta^2, \quad x_2^{(2)} = 9\delta^2/\pi - i\delta^2. \tag{19}$$

The presence of the imaginary part indicates damping in view of possible departure of an electron from the

TWNS. Relations (17)–(19) make it possible to write determinant $\Delta_0(2)$ in the form

$$\Delta_0(2) \approx \frac{8i}{\Gamma t_0}[\varepsilon - \varepsilon_R^{(1)} + i\Gamma][\varepsilon - \varepsilon_R^{(2)} + i\Gamma], \quad \Gamma = \frac{2p_0^3}{\alpha^2 a}, \tag{20}$$

where Γ is the width of the resonance levels in the TWNS. It can easily be seen that the determinant assumes the following values in resonance:

$$\Delta_0(2, p = p^{(1)}) = -8, \quad \Delta_0(2, p = p^{(2)}) = 8. \tag{21}$$

Expression (20) makes it possible to find the TWNS transmission coefficient in explicit form:

$$T(\varepsilon) = \frac{\Gamma^2 t_0^2}{[(\varepsilon - \varepsilon_R^{(1)})^2 + \Gamma^2][(\varepsilon - \varepsilon_R^{(2)})^2 + \Gamma^2]}. \tag{22}$$

This coefficient has two resonance peaks in the vicinity of resonance energies $\varepsilon_R^{(1,2)}$ with width Γ. The resonant values $T(\varepsilon_R^{(1,2)})$ are equal to unity, while for $\varepsilon = \varepsilon_R^{(1)} - t_0/2$ (the value between the maxima), the transmission coefficient becomes quite small, $T(\varepsilon_R^{(1)} - t_0/2) = 4(p/\alpha)^2 \ll 1$.

Let us consider the general case $V \neq 0$, assuming, however, that $V \ll \varepsilon_R$. In this case, the equations are extremely cumbersome and will not be given here in complete form. In the first approximation in δ, the equation for x_1 assumes the form

$$x_1^2 + 2x_1\left(2\delta + \frac{\tilde{V}}{4}\right) + 3\delta^2 + \delta\tilde{V} = 0, \tag{23}$$

where $\tilde{V} = Va/p_0 \ll 1$ and $\delta = p_0/\alpha \ll 1$.

Equation (23) has two solutions,

$$x_1^{(1,2)} = -\left(2\delta + \frac{\tilde{V}}{4}\right) \pm \sqrt{\delta^2 + \left(\frac{\tilde{V}}{4}\right)^2}, \tag{24}$$

which correspond to two levels with the corresponding energies

$$\varepsilon_R^{(1,2)}(1) = p_0^2 + \frac{2p_0 x_1^{(1,2)}}{a}. \tag{25}$$

The value of energy splitting

$$t = \varepsilon_R^{(1)}(1) - \varepsilon_R^{(2)}(2) = t_0\sqrt{1 + V^2/t_0^2} \tag{26}$$

increases with V. For $V \gg t_0$, the splitting is $t \approx V$; i.e., the energy levels in the wells become independent.

The imaginary part of the correction of the next order in δ and $\tilde{\varepsilon}_1$ has the universal form

$$x_2'' = -\delta^2 \tag{27}$$

for both levels and is independent of $\tilde{V}$. Consequently, the level widths Γ are identical and are given by expression (20). The real part of correction x_2' is cumbersome and is not given here. Whenever necessary, it can be taken into account in $\varepsilon_R^{(1,2)}$ and t.

Substituting the values of resonance energy (to within δ^2, $\tilde{V}^2$, and $\delta\tilde{V}$) into determinant $\Delta_{10}(2)$, we obtain the expression

$$\Delta_{10}(2, p = p^{(1,2)}) = \mp 8\sqrt{1 + V^2/t_0^2}, \tag{28}$$

which is transformed into (21) for $V = 0$. It can be seen that the resonant value increases with V as expected. Using the results obtained above, we can represent determinant $\Delta_{10}(2)$ in the vicinity of the resonance in the form

$$\Delta_{10}(2) = \frac{i8}{\Gamma t_0}[\varepsilon - \varepsilon_R^{(1)}(1) + i\Gamma][\varepsilon - \varepsilon_R^{(2)}(1) + i\Gamma]. \tag{29}$$

As before, transmission coefficient (12),

$$T(\varepsilon) = \frac{\Gamma^2 t_0^2}{[(\varepsilon - \varepsilon_R^{(1)}(1))^2 + \Gamma^2][(\varepsilon - \varepsilon_R^{(2)}(1))^2 + \Gamma^2]}, \tag{30}$$

has two resonance peaks of width Γ. The resonant values of $T(\varepsilon)$ decrease with increasing V,

$$T(\varepsilon) = \frac{t_0^2}{t_0^2 + V^2}, \tag{31}$$

since field V leads to degradation of interwell interference and to localization of electrons in individual wells.

4. WAVEFUNCTIONS IN A TWO-WELL STRUCTURE IN THE PRESENCE OF A VARYING FIELD: GENERAL SOLUTION

The solutions to Eqs. (6) for functions Ψ_{ln} have the form (see, for example, [7])

$$\Psi_{1n} = A_{1n}e^{ip_n x} + B_{1n}e^{-ip_n x} + f_1(x),$$
$$\Psi_{2n} = A_{2n}e^{ip_{1n} x} + B_{2n}e^{-ip_{1n} x} + f_2(x), \tag{32}$$
$$f_l(x) = -\frac{2U}{\omega_n^2}\left[\Psi_{l0}'(x) + \frac{x\omega_n}{2}\Psi_{l0}(x)\right].$$

Substituting Ψ_{1n} and Ψ_{2n} into boundary conditions (7), we obtain the following system of linear equations for determining coefficients A_{ln} and B_{ln}:

$$\begin{aligned}
&(2-\beta_n)A_{1n}-\beta_n B_{1n} = F_1,\\
&A_{1n}+B_{1n}e^{-2ip_na}-A_{2n}e^{i(p_{1n}-p_n)a}\\
&\quad -B_{2n}e^{-i(p_{1n}+p_n)a} = F_2,\\
&A_{1n}(\beta_n+1)+B_{1n}(\beta_n+1)\\
&-\frac{p_{1n}}{p_n}e^{i(p_{1n}-p_n)a}A_{2n}+\frac{p_{1n}}{p_n}e^{-i(p_{1n}+p_n)a}B_{2n} = F_3,\\
&\beta_{1n}A_{2n}+(\beta_{1n}-2)B_{2n}e^{-4ip_{1n}a} = F_4.
\end{aligned} \tag{33}$$

The right-hand sides F_j are given by

$$\begin{aligned}
F_1 &= (\beta_n-1)f_1(0)-\frac{f_1'(0)}{ip_n},\\
F_2 &= [f_2(a)-f_1(a)]e^{-ip_na},\\
F_3 &= \left[\frac{f_2'(a)}{ip_n}-\frac{f_1'(a)}{ip_n}-\beta_n f_1(a)\right]e^{-ip_na},\\
F_4 &= \left[(1-\beta_{1n})f_2(2a)-\frac{f_2'(2a)}{ip_{1n}}\right]e^{-2ip_{1n}a}.
\end{aligned} \tag{34}$$

The solutions to system (33) can be represented in the form

$$\begin{aligned}
A_{1n}\Delta_{1n}(2) &= F_1\Delta_{1n}(1)e^{2ip_{1n}a}+\beta_n\{\mathrm{I}\},\\
B_{1n}\Delta_{1n}(2) &= F_1\Pi_{1n}e^{2ip_na}+(2-\beta_n)\{\mathrm{I}\},\\
A_{2n}\tilde{\Delta}_{1n}(2) &= -F_4\tilde{\Pi}_{1n}e^{-i(p_{1n}+p_n)a-2ip_na}\\
&\quad +(2-\beta_n)e^{-4ip_{1n}a}\{\mathrm{II}\},\\
B_{2n}\tilde{\Delta}_{1n}(2) &= -F_4\tilde{\Delta}_{1n}(1)e^{i(p_{1n}-p_n)a}+\beta_{1n}\{\mathrm{II}\},
\end{aligned} \tag{35}$$

where the following notation is introduced:

$$\begin{aligned}
\{\mathrm{I}\} &= e^{2ip_na}\left[\frac{p_{1n}}{p_n}F_2C_-^n-F_3C_+^n\right]\\
&\quad -2\frac{p_{1n}}{p_n}F_4\exp[2ip_{1n}a+i(p_{1n}+p_n)a],\\
\{\mathrm{II}\} &= 2F_1e^{-2ip_na}+[F_2D_n-F_3C_n],
\end{aligned} \tag{36}$$

$$\begin{aligned}
\tilde{\Delta}_{1n}(2) &= \exp[-3i(p_{1n}+p_n)a]\Delta_{1n}(2),\\
\Delta_{1n}(2) &= (2-\beta_n)\Delta_{1n}(1)e^{2ip_{1n}a}-\beta_n\Pi_{1n}e^{2ip_na},
\end{aligned} \tag{37}$$

$$\begin{cases}
\Delta_{1n}(1) = (2-\beta_{1n})\left(1+\dfrac{p_{1n}}{p_n}-\beta_n\right)e^{-2ip_{1n}a}\\
\quad -\beta_{1n}\left(\beta_n+\dfrac{p_{1n}}{p_n}-1\right),\\
\Pi_{1n}=\left(1+\beta_n-\dfrac{p_{1n}}{p_n}\right)(1-\beta_{1n})+\beta_{1n}\left(1+\dfrac{p_{1n}}{p_n}+\beta_n\right)e^{2ip_{1n}a},\\
C_\pm^n = (2-\beta_{1n})\pm\beta_{1n}e^{2ip_{1n}a},\\
C_n = (2-\beta_n)e^{-2ip_na}+\beta_n,\\
D_n = \beta_n(\beta_n+1)+(\beta_n-1)(2-\beta_n)e^{-2ip_na},\\
\tilde{\Delta}_{1n}(1) = (2-\beta_n)\left(1+\dfrac{p_{1n}}{p_n}-\beta_n\right)e^{-2ip_na}\\
\quad -\beta_n\left(\beta_n+1-\dfrac{p_{1n}}{p_n}\right),\\
\tilde{\Pi}_{1n}=(2-\beta_n)\left(\beta_n+\dfrac{p_{1n}}{p_n}-1\right)+\beta_n\left(\beta_n+1+\dfrac{p_{1n}}{p_n}\right)e^{2ip_na},\\
\Delta_{1n}(2) = (1-\beta_{1n})\tilde{\Delta}_{1n}(1)e^{2ip_na}-\beta_{1n}\tilde{\Pi}_{1n}e^{2ip_{1n}a}.
\end{cases} \tag{38}$$

Determinant $\Delta_{1n}(2)$ in (38) is analogous to $\Delta_{10}(2)$ to within the substitution $p_{1n}\longrightarrow p_1$. Consequently, it can be written in the form

$$\begin{aligned}
\Delta_{1n}(2) &\approx \frac{i8}{\Gamma t_0}[\varepsilon-\omega_n-\varepsilon_R^{(1)}(1)+i\Gamma]\\
&\quad \times[\varepsilon-\omega_n-\varepsilon_R^{(2)}(1)+i\Gamma].
\end{aligned} \tag{39}$$

Formulas (35)–(39) provide an exact solution to system (6), (7), which is unfortunately cumbersome and hard to visualize. In addition, this solution contains formally diverging expressions for $\omega\longrightarrow 0$. However, as was shown for the first time in [7], there exists a transformation under which diverging terms can be singled out and compensated exactly. In this case, use is made of small parameters $\delta = p_0/\alpha$ and ω_n/ε_R, which are present in a structure with high barriers. The smallness of the frequency (equal to splitting t in resonance) as compared to the resonance energy follows from the smallness of splitting,

$$\frac{t}{\varepsilon_R} = \frac{4}{\pi}\delta \ll 1.$$

In particular, on account of the smallness of δ and ω_n/ε_R, we can simplify the expressions for F_j, omitting

the terms which are small in δ and ω_n/ε_R:

$$F_1 = F_{11} = \bar{U}[(2-\beta_n)A_{10} + \beta_n B_{10}],$$
$$\bar{U} = \frac{2Uip}{\omega_n^2}, \tag{40}$$
$$F_2 = F_{21} + F_{22}, \quad F_3 = F_{31} + F_{32},$$

$$F_{21} = \bar{U}\left\{[A_{10} - B_{10}e^{-2ipa}] - i\frac{a\omega_n}{2p}(A_{10} + B_{10}e^{-2ipa})\right\}e^{i(p-p_n)a},$$

$$F_{31} = -\frac{\bar{U}p_1}{p}\left\{[A_{10}(1+\beta_n) + B_{10}e^{-2ipa}(1-\beta_n)] - i\frac{a\omega_n}{2p}[A_{10}(1+\beta_n) + B_{10}e^{-2ipa}(\beta_n - 1)]\right\}e^{i(p-p_n)a},$$

$$F_{22} = -\frac{\bar{U}p_1}{p}\left\{[A_{20} - B_{20}e^{-2ip_1a}] - i\frac{a\omega_n}{2p_1}[A_{20} + B_{20}e^{-2ip_1a}]\right\}\exp(i(p_1 - p_n)a), \tag{41}$$

$$F_{32} = -\frac{\bar{U}p_1p_{1n}}{pp_n}\left\{[A_{20} + B_{20}e^{-2ip_1a}] - i\frac{a\omega_n}{2p}[A_{20} - B_{20}e^{-2ip_1a}]\right\}\exp(i(p_1 - p_n)a),$$

$$F_{42} = \bar{U}\frac{p_1}{p}\{\beta_{1n}A_{20} + (2-\beta_{1n})B_{20}e^{-4ip_1a}\} \times \exp(2i(p_1 - p_{1n})a).$$

5. WAVEFUNCTION AND RESPONSE OF THE FIRST WELL IN A TWO-WELL STRUCTURE

It is more convenient to consider the wavefunctions and the response of each well separately. First, we express coefficients A_{1n} and B_{1n} in terms of A_{l0} and B_{l0} so as to single out the terms diverging for $\omega \longrightarrow 0$,

$$A_{1n} = \bar{U}A_{10} - \frac{\bar{U}\beta_n}{\Delta_{1n}(2)}[M_{1n} - \Phi_{1n}] \equiv A_{1n}^{(1)} + A_{1n}^{(2)},$$
$$B_{1n} = \bar{U}B_{10} - \frac{\bar{U}(2-\beta_n)}{\Delta_{1n}(2)}[M_{1n} - \Phi_{1n}] \equiv B_{1n}^{(1)} + B_{1n}^{(2)}, \tag{42}$$

where the following notation is introduced:

$$M_{1n} = [A_{10}\Pi_{1n}Y_n e^{2ip_na} + B_{10}\Delta_{1n}(1)Y_n^* e^{2ip_{1n}a}],$$

$$\Phi_{1n} = 2\frac{p_1p_{1n}}{pp_n}\exp(i(p_{1n} + p_n)a) \times [A_{20}B_{1n}X_{1n}^* e^{2ip_1a} + B_{20}(2-\beta_n)X_{1n}e^{-2ip_1a}],$$

$$Y_n = Z_n - i\frac{a\omega_n}{2p}(1 + Z_n), \tag{43}$$

$$X_{1n} = Z_{1n} + i\frac{a\omega_n}{2p}(1 + Z_{1n}),$$
$$Z_n = e^{i(p-p_n)a} - 1, \quad Z_{1n} = e^{i(p_1-p_{1n})a} - 1.$$

Contributions $A_{1n}^{(1)}$ and $B_{1n}^{(1)}$ contain terms diverging for $\omega \longrightarrow 0$ and are independent of determinant $\Delta_{1n}(2)$, which is cancelled out due to the terms grouped in $\Delta_{1n}(2)$ when A_{10} and B_{10} appear in the numerator. Coefficients $A_{1n}^{(2)}$ and $B_{1n}^{(2)}$ are finite for $\omega \longrightarrow 0$ and are the only terms contributing to the response.

It can be proved that the part of the wavefunction diverging in ω and emerging due to $A_{1n}^{(1)}$ and $B_{1n}^{(1)}$ is compensated by the contribution from the inhomogeneous solution f_1. Indeed, expanding $p_{1n} = p_1 - \omega_n/2p_1$ in the exponent $\exp(\pm ip_{1n}x)$ appearing with $A_{1n}^{(1)}$ and $B_{1n}^{(2)}$, we see that the corresponding diverging contributions are cancelled out. Thus, for Ψ_{1n} we obtain

$$\Psi_{1n} \approx A_{1n}^{(2)}e^{ip_{1n}x} + B_{1n}^{(2)}e^{-ip_{1n}x}. \tag{44}$$

It will be shown below that current (8) can be expressed in terms of combinations of $A_{1n}^{(2)}$ and $B_{1n}^{(2)}$,

$$\delta_{1n} = A_{1n}^{(2)} - B_{1n}^{(2)}, \quad \gamma_{1n} = A_{1n}^{(2)} + B_{1n}^{(2)}, \tag{45}$$

which can be determined from relations (42):

$$\gamma_{1n} = -\frac{2\bar{U}}{\Delta_{1n}(2)}[M_{1n} - \Phi_{1n}],$$
$$\delta_{1n} = (\beta_{1n} - 1)\gamma_{1n}. \tag{46}$$

It should be noted that the coefficients of wavefunction Ψ_{1n} and their combinations γ_{1n} and δ_{1n} contain the interfering contribution both from the electrons arriving from the upper level of the first well after the emission of a photon (on the order of M_{1n}) and from the electrons

experiencing resonant tunnelling from the second well (on the order of Φ_{1n}).

Let us now calculate M_{1n} using expressions (43), (10), and (11). In the main approximation in δ and $\tilde{V}$, determinant $\Delta_{10}(1)$ for the resonance value of momentum $p = p^{(1)}$ is given by

$$\Delta_{10}(1) \approx -\Pi_{10}(p^{(1)}) \approx \frac{2i\alpha}{p_0}[V_0 + \sqrt{1 + V_0^2}], \quad V_0 = \frac{V}{t_0}. \tag{47}$$

Here and below, we assume that energy ε of the electrons supplied to the structure is equal to $\varepsilon_R^{(1)}(1)$. To determine the value of $\Delta_{1n}(1)$, we note that it coincides with $\Delta_{10}(1)$ if we replace p_{1n} by p_1 and p_n by p. Consequently, choosing the frequency equal to splitting energy t, we find that $p_{1n}^{(1)} = p_1^{(2)}$, and hence

$$\Delta_{1n}(1, p^{(1)}) \approx -\Pi_{1n} = \frac{2i\alpha}{p_0}[V_0 - \sqrt{1 + V_0^2}]. \tag{48}$$

Substituting cofactors into Eq. (43) and considering that

$$Y_n + Y_n^* = \left(\frac{a\omega_n}{2p}\right)^2, \tag{49}$$

we obtain

$$M_{1n} = -\frac{8q\alpha^2}{\Delta_{10}(2)p^2}\left(\frac{a\omega_n}{2p}\right)^2. \tag{50}$$

It can easily be seen that

$$\Phi_{1n} = 3M_{1n}, \tag{51}$$

since

$$X_{1n} + X_{1n}^* = -3\left(\frac{a\omega_n}{2p}\right)^2. \tag{52}$$

Summing the results, we obtain

$$\gamma_{1n} = -\frac{i4Ua(8q)}{\Gamma\Delta_{10}(2)\Delta_{1n}(2)}. \tag{53}$$

It should be emphasized that the squared frequency in the denominator cancels out; consequently, function Ψ_{1n} is finite for $\omega \longrightarrow 0$.

Let us now calculate the current $J_{1n}^c(x)$ in the first well. Substituting functions Ψ_{10} and Ψ_{1n} into Eq. (8), we arrive at the following expression:

$$J_{1n}^c(x) = ep\{[K_{1n} + \text{c.c.}]\cos(p - p_n)x - i[F_{1n} - \text{c.c.}]\sin(p - p_n)x\}, \tag{54}$$

$$K_{1n} = (\delta_{10}^*\gamma_{1n} + \gamma_{10}^*\delta_{1n}), \quad p - p_n \approx \omega_n/2p,$$
$$F_{1n} = (\delta_{10}^*\delta_{1n} + \gamma_{10}^*\gamma_{1n}), \quad \delta_{10} = A_{10} - B_{10}, \quad \gamma_{10} = A_{10} + B_{10}. \tag{55}$$

Current J_{1n}^c (54) is the sum of two terms. The first term weakly depends on the coordinate and is proportional to φ_{1n},

$$K_{1n} = \gamma_{1n}\varphi_{1n}, \tag{56}$$

$$\varphi_{1n} = \delta_{10}^* + (\beta_n - 1)\gamma_{10}^* = \beta_n(A_{10}^* + B_{10}^*) - 2B_{10}^*. \tag{57}$$

It was shown in [7] for a single-well structure that the first term describes electron transitions between states with wavefunctions proportional to $\sin p_n x$ and $\cos px$, $\sin px$ and $\cos p_n x$, which are typical of the current state and vanish in an isolated quantum well. These transitions occur with approximately equal weights, thus making intense interference of these states possible. If the energy of supplied electrons coincides with the resonance energy, compensation takes place and φ_{1n} vanishes. It will be shown below that, as before, the contribution of the φ term is also small for any V if $\varepsilon = \varepsilon_R^{(1)}(1)$.

The second term corresponds to a "laser-type" transition between states $\sin px$ and $\sin p_n x$ since these functions coincide with the eigenfunctions of an isolated well. True, momenta p_1 and p_n differ by a small quantity $\omega_n/2p$ (in a laser, $p - p_n = \pm\pi/a$).

Omitting terms small in δ, we obtain the following expression for F_{1n}:

$$\{F_{1n} - \text{c.c.}\} = i\frac{(8q)^2 4Ua\alpha^2/p^2[V_0 + \sqrt{1 + V_0^2}]}{\Gamma|\Delta_{10}(2)|^2} \times\left[\frac{1}{\Delta_{1n}(2)} + \text{c.c.}\right]. \tag{58}$$

It follows hence that current J_{1n}^c is proportional to $[1/\Delta_{1n}(2) + \text{c.c.}]$. It follows from expression (39) for $\Delta_{1n}(2)$ that $\Delta_{1,1}(2) \gg \Delta_{1,-1}$ since the resonance conditions are satisfied for a transition between levels with energies $\varepsilon_R^{(1)}$ and $\varepsilon_R^{(2)}$. For this reason, we will hence-

forth disregard nonresonant transitions, assuming that $J_{1c}(x) \approx J^c_{1,-1}(x)$.

We can now find the reduced current and the response of the first well:

$$J_{1c} = \int_0^a J^c_{1,-1}(x)\frac{dx}{a},$$
$$\tilde{J}_{1c} = \frac{J_{1c}}{(Ua)(epq^2)} = \frac{\omega_n a}{4p} \tag{59}$$
$$\times \frac{(8)^2 4\alpha^2/p^2[V_0 + \sqrt{1+V_0^2}]}{\Gamma|\Delta_{10}(2)|^2}\left[\frac{1}{\Delta_{1n}(2)} + \text{c.c.}\right].$$

Assuming that the resonance condition $\omega_n = t = t_0\sqrt{1+V_0^2}$ is satisfied and using the resonance values of $\Delta_{10}(2)$ and $\Delta_{1n}(2)$, we arrive at the final formula

$$\tilde{J}_{1c}(V_0) = \tilde{J}_{1c}(0)\frac{[V_0 + \sqrt{1+V_0^2}]}{1+V_0^2}, \tag{60}$$
$$\tilde{J}_{1c}(0) = \frac{\alpha}{\Gamma p}.$$

Let us analyze the dependence of J_{1c} on the bias voltage V_0. For $V_0 = 0$, the current assumes the value $J_{1c}(0)$, then increases with V_0, and decreases in proportion to $1/V_0$ for large values of $V_0 \gg 1$. For a certain value of $V_0 = 1/\sqrt{3}$, the current attains its maximal value. The increase in current J_{1c} in a certain interval of bias voltages V can be explained by the fact that field V "detunes" the resonant determinant $\Delta_{10}(1)$, increasing it in proportion to $[V_0 + \sqrt{1+V_0^2}]$. As the value of V increases further, the increase in $|\Delta_{10}(2)|^2$ and $\Delta_{1n}(2)$ becomes predominant, and J_{1c} attains is maximal value. It can be concluded hence that, in a certain optimal interval of variation $0 < V \le t_0$, the gain is maximal and frequency tuning is possible.

Concluding this section, we will prove that if the energy of supplied electron is equal to the energy of the upper level, the value of φ_{1n} is small as compared to the second term F_{1n} in parameter δ. Indeed, substituting A_{10} and B_{10} into relation (57), we can reduce the expression for φ_{1n} to the form

$$\varphi_{1n} = -\frac{2q}{\Delta_{10}(2)}$$
$$\times\{2[e^{2ipa}\Pi_{10} + e^{2ip_1a}\Delta_{10}(1)] - \Delta_{10}(2)\}$$
$$\approx \frac{2iq8}{\Delta_{10}(2)}\{2V_0[V_0 + \sqrt{1+V_0^2}] - \sqrt{1+V_0^2}\}.$$

This readily leads to the above statement.

6. WAVEFUNCTION AND RESPONSE OF THE SECOND WELL

Proceeding in the same way as for the first well, we can determine coefficients A_{2n} and B_{2n} by separating the terms diverging in ω. After certain transformations, we obtain

$$A_{2n} = \bar{U}_1 A_{20}\exp(2i(p_1 - p_{1n})a)$$
$$+ \bar{U}\frac{2-\beta_{1n}}{\Delta_{1n}(2)}(M_{2n} - \Phi_{2n}),$$
$$B_{2n} = -\bar{U}_1 B_{20}\exp(-2i(p_1 - p_{1n})a) \tag{61}$$
$$+ \bar{U}\frac{\beta_{1n}e^{4ip_{1n}a}}{\Delta_{1n}(2)}(M_{2n} - \Phi_{2n}),$$

where

$$M_{2n} = [A_{20}e^{2ip_1a}\tilde{\Delta}_{1n}(1)X^*_{1n}\exp(2i(p_n - p_{1n})a)$$
$$+ B_{20}e^{-2ip_1a}\tilde{\Pi}_{1n}X_{1n}],$$
$$\Phi_{2n} = 2\exp(2i(p_n - p_{1n})a) \tag{62}$$
$$\times[A_{10}(2-\beta_n)Y_n + B_{10}\beta_n Y^*_n], \quad \bar{U}_1 = \bar{U}\frac{p_1}{p}.$$

Expressions (61) for A_{2n} and B_{2n} differ from formula (42) for A_{1n} and B_{1n} in that the terms with A_{20} and B_{20} contain exponentials $\exp(\pm 2i(p_1 - p_{1n})a)$ and $e^{4ip_{1n}a}$ in the second term for B_{2n}. To restore the symmetry, it is convenient to pass to new functions $\tilde{\psi}_{2n}$ and $\tilde{\psi}_{20}$:

$$\tilde{\psi}_{2n} = \tilde{A}_{2n}\exp(ip_n(x-2a))$$
$$+ \tilde{B}_{2n}\exp(-ip_n(x-2a)) + f_2(x-2a),$$
$$\tilde{\psi}_{20} = \tilde{A}_{20}\exp(ip_1(x-2a)) \tag{63}$$
$$+ \tilde{B}_{20}\exp(-ip_1(x-2a)),$$

where

$$\tilde{A}_{2n} = A_{2n}e^{2ip_{1n}a}, \quad \tilde{B}_{2n} = B_{2n}e^{-2ip_{1n}a},$$
$$\tilde{A}_{20} = A_{20}e^{2ip_1a}, \quad \tilde{B}_{20} = B_{20}e^{-2ip_1a}. \tag{64}$$

After the transformation of expressions (63) and (64),

the expressions for $\tilde{A}_{2n}$ and $\tilde{B}_{2n}$ assume the form

$$\begin{aligned}\tilde{A}_{2n} &= \overline{U}_1\tilde{A}_{20} + \frac{\overline{U}(2-\beta_{1n})}{\Delta_{1n}(2)}[\tilde{M}_{2n} - \tilde{\Phi}_{2n}] \\ &\equiv \tilde{A}_{2n}^{(1)} + \tilde{A}_{2n}^{(2)}, \\ \tilde{B}_{2n} &= -\overline{U}_1\tilde{B}_{20} + \frac{\overline{U}\beta_{1n}}{\Delta_{1n}(2)}[\tilde{M}_{2n} - \tilde{\Phi}_{2n}] \\ &\equiv \tilde{B}_{2n}^{(1)} + \tilde{B}_{2n}^{(2)},\end{aligned} \tag{65}$$

where

$$\begin{aligned}\tilde{M}_{2n} &= [\tilde{A}_{20}\tilde{\Delta}_{1n}(1)e^{2ip_n a}X_{1n}^* + \tilde{B}_{20}\tilde{\Pi}_{1n}e^{2ip_1 a}X_{1n}], \\ \tilde{\Phi}_{2n} &= 2e^{i(p_{1n}+p_n)a}[A_{10}(2-\beta_n)Y_n + B_{10}\beta_n Y_n^*].\end{aligned} \tag{66}$$

In addition, the expression for the current in the second well is transformed to

$$\begin{aligned}J_{2n}^c(x) &= ep\{(\tilde{K}_2 + \text{c.c.})\cos(p_1 - p_{1n})(x-2a) \\ &\quad - i(\tilde{F}_{2n} - \text{c.c.})\sin(p_1 - p_{1n})(x-2a)\}, \\ p_1 - p_{1n} &\approx \frac{\omega_n}{2p_1}.\end{aligned} \tag{67}$$

In expressions for $\tilde{K}_{2n}$ and $\tilde{F}_{2n}$, we must replace coefficients A and B by $\tilde{A}$ and $\tilde{B}$. Using relations (65), we obtain the following expressions for $\tilde{\gamma}_{2n}$ and $\tilde{\delta}_{2n}$:

$$\begin{aligned}\tilde{\gamma}_{2n} &= \tilde{A}_{2n}^{(2)} + \tilde{B}_{2n}^{(2)} = \frac{2\overline{U}}{\Delta_{1n}(2)}[\tilde{M}_{2n} - \tilde{\Phi}_{2n}] \\ \tilde{\delta}_{2n} &= \tilde{A}_{2n}^{(2)} - \tilde{B}_{2n}^{(2)} = -(\beta_{1n}-1)\tilde{\gamma}_{2n}.\end{aligned} \tag{68}$$

It should be noted that the relation between $\tilde{\delta}_{2n}$ and $\tilde{\gamma}_{2n}$ has the sign opposite to relation (46); consequently the structure of $\tilde{\varphi}_{2n}$ differs from that of φ_{1n}. First, we will calculate $\tilde{K}_{2n}$:

$$\tilde{K}_{2n} = \tilde{\gamma}_{2n}\tilde{\varphi}_{2n}, \quad \tilde{\varphi}_{2n} = 2\tilde{A}_{20}^* - \beta_{1n}(\tilde{A}_{20}^* + \tilde{B}_{20}^*). \tag{69}$$

Substituting $\tilde{A}_{20}$ and $\tilde{B}_{20}$, we obtain

$$\begin{aligned}\tilde{\varphi}_{2n} &= \frac{8qe^{-i(p_1+p)a}}{\Delta_{10}^*(2)}[2 + (\beta_{1n} - \beta_1)] \\ &= \frac{16qe^{-i(p_1+p)a}}{\Delta_{10}^*(2)}\left[1 + \sqrt{1 + \frac{\varepsilon_1^2}{t_0^2}}\right].\end{aligned}$$

As in the case of the first well, the φ contribution is small in parameter α/p as compared to the "laser" contribution from $\tilde{F}_{2n}$. Thus, we observe that TWNSs differ basically from RTDs, for which the φ contribution dominates.

Let us find the reduced current (67) in the second well:

$$\begin{aligned}J_{2n}^c &= \frac{1}{a}\int_a^{2a} dx\sin(p_1 - p_{1n})(x-2a) \\ &\times [-iep_1(\tilde{F}_{2n} - \text{c.c.})] = \left(-\frac{\omega_n a}{4p_1}\right)[-iep_1(\tilde{F}_{2n} + \text{c.c.})].\end{aligned} \tag{70}$$

It should be observed that the integral is negative in contrast to the integral for the first well. However, $\tilde{\gamma}_{2n}$ also has the opposite sign as compared to γ_{1n}.

We will now calculate $\tilde{M}_{2n}$, which involves certain difficulties. We will proceed as follows. We decompose X_{1n} and X_{1n}^* into the real and imaginary parts,

$$\begin{aligned}X_{1n} &= X + iY, \quad X_{1n}^* = X - iY, \\ X &= -\frac{3}{2}\left(\frac{a\omega_n}{2p_1}\right)^2, \quad Y = 2\left(\frac{a\omega_n}{2p_1}\right).\end{aligned} \tag{71}$$

Now, we can write the expression for $\tilde{M}_{2n}$ in the form

$$\begin{aligned}\tilde{M}_{2n} &= X\{\tilde{A}_{20}\tilde{\Delta}_{1n}(1)e^{2ip_n a} + \tilde{B}_{20}\tilde{\Pi}_{1n}e^{2ip_{1n}a}\} \\ &\quad + iY\{\tilde{B}_{20}\tilde{\Pi}_{1n}e^{2ip_{1n}a} - \tilde{A}_{20}\tilde{\Delta}_{1n}(1)e^{2ip_n a}\}.\end{aligned}$$

The term containing Y can be reduced to determinant $\Delta_{1n}(2)$ whose resonance value is small. Consequently, the second term can be disregarded. It can easily be proved that the expression for $\tilde{\Pi}_{1n}$ for $p = p^{(1)}$ has the form

$$\begin{aligned}\tilde{\Pi}_{1n}(p = p^{(1)}) &= -\tilde{\Delta}_{1n}(1) \\ &= \frac{2i\alpha}{p}\left[\frac{V}{t_0} + \sqrt{1 + \frac{V^2}{t_0^2}}\right],\end{aligned} \tag{72}$$

so that we can write

$$\tilde{M}_{2n} = -3\left(\frac{a\omega_n}{2p_1}\right)^2\frac{8q\alpha^2/p^2}{\Delta_{10}(2)}[V_0 + \sqrt{1+V_0^2}]. \tag{73}$$

Accordingly, for $\tilde{\Phi}_{2n}$ and $\tilde{\gamma}_{2n}$ we obtain

$$\tilde{\Phi}_{2n} = \frac{1}{3}\tilde{M}_{2n}, \tag{74}$$

$$\tilde{\gamma}_{2n} = -\frac{i(8q)4aU}{\Delta_{10}(2)\Delta_{1n}(2)\Gamma}[V_0 + \sqrt{1+V_0^2}]. \tag{75}$$

Substituting $\tilde{\delta}_{20}^{*}$ and $\tilde{\delta}_{2n} \approx -\beta\tilde{\gamma}_{2n}$ into expression (70), we obtain the current in the second well,

$$\tilde{J}_{2n}^{c} = \frac{(8)^2 4\alpha/p}{|\Delta_{10}(2)|^2\Gamma}\left(\frac{1}{\Delta_{1n}} + \text{c.c.}\right) \times [V_0 + \sqrt{1+V_0^2}]\sqrt{1+V_0^2}, \quad (76)$$

which completely coincides with current (59) in the first well. The corresponding resonance value has the form

$$\tilde{J}_{2c} = \tilde{J}_c(0)\left[\frac{V_0 + \sqrt{1+V_0^2}}{1+V_0^2}\right], \quad \tilde{J}_c(0) = \frac{\alpha/p}{\Gamma}. \quad (77)$$

Equation (2) was also solved numerically [10]. The analytic results obtained above for the transmission coefficients of the currents in the first and second wells coincide with numerical values.

7. DISCUSSION OF RESULTS AND CONCLUSIONS

It would be interesting to compare the results obtained for a TWNS with the corresponding results for a single-well structure (RTD) and for a coherent quantum-well laser. According to [7], the polarization current (response) of an RTD is given by

$$\tilde{J}_c = \frac{J_c}{Uapq^2} = \frac{4\gamma\Gamma_1^2}{[(\omega-\gamma)^2+\Gamma_1^2][(\omega+\gamma)^2+\Gamma_1^2]}, \quad (78)$$
$$\gamma = \varepsilon - \varepsilon_R, \quad \Gamma_1 = 2\Gamma.$$

In the low-frequency limit, we have $\omega \ll \Gamma$; assuming that $\gamma = \Gamma_1$, we obtain $\tilde{J}_c = 1/\Gamma_1$. Comparing (78) with expression (77), we see that the reduced high-frequency current ($\omega \gg \Gamma$) in the TWNS exceeds the current in the RTD by a factor larger than α/p. A still larger gain as compared to the RTD current is obtained in the high-frequency quantum regime [7] for $\gamma = \omega \gg \Gamma$:

$$\tilde{J}_c = \frac{1}{\omega}. \quad (79)$$

The current ratio is

$$\frac{J_{1c}}{J_c} \approx \frac{\alpha}{p}\frac{\omega}{\Gamma}. \quad (80)$$

Thus, a TWNS provides a substantially higher gain at high frequencies $\omega \gg \Gamma$ as compared to a single-well structure.

Comparing J_{1c} with the corresponding current in a coherent laser [2], in which radiative transitions occur between the resonance levels in a quantum well ($p - p_n = \pi/a$), we find that the value of J_{1c} is smaller by p/α. The reason is that the frequency is lower and equal to the splitting energy. However, in contrast to a coherent laser, the frequency of the TWNS can be tuned by changing the constant voltage V, the voltage increasing the gain in a certain interval of V.

ACKNOWLEDGMENTS

The author thanks Yu.V. Kopaev and I.Yu. Kateev for fruitful discussions.

This study was supported by Federal target program "Integration" (project no. B0049) and by the program "Physics of Solid Nanostructures" of the Ministry of Industry and Science of the Russian Federation (grant no. 99-1140).

REFERENCES

1. J. P. Sun, G. I. Haddad, P. Mazumder, and J. N. Schulman, Proc. IEEE **86**, 644 (1998).
2. V. F. Elesin, Zh. Éksp. Teor. Fiz. **112**, 483 (1997) [JETP **85**, 264 (1997)].
3. V. F. Elesin and Yu. V. Kopaev, Zh. Éksp. Teor. Fiz. **123**, 1308 (2003) [JETP **96**, 1149 (2003)].
4. S. Haas, T. Stroucken, M. Hübner, *et al.*, Phys. Rev. B **57**, 14860 (1998).
5. I. A. Dmitriev and R. A. Suris, Fiz. Tekh. Poluprovodn. (St. Petersburg) **36**, 1460 (2002) [Semiconductors **36**, 1375 (2002)].
6. V. F. Elesin, JETP **96**, 966 (2003).
7. V. F. Elesin, Zh. Éksp. Teor. Fiz. **116**, 704 (1999) [JETP **89**, 377 (1999)]; **121**, 925 (2002) [**94**, 794 (2002)]; **124**, 379 (2003) [**97**, 343 (2003)].
8. V. F. Elesin, V. V. Kapaev, Yu. V. Kopaev, and A. V. Tsukanov, Pis'ma Zh. Éksp. Teor. Fiz. **66**, 709 (1997) [JETP Lett. **66**, 742 (1997)].
9. E. I. Golant and A. B. Pashkovskiĭ, Fiz. Tekh. Poluprovodn. (St. Petersburg) **36**, 330 (2002) [Semiconductors **36**, 311 (2002)].
10. V. F. Elesin and I. Yu. Kateev, Fiz. Tekh. Poluprovodn. (St. Petersburg) (in press).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 100, No. 1, 2005, pp. 126–138.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 1, 2005, pp. 141–155.

ELECTRONIC PROPERTIES OF SOLIDS

Energy Relaxation of Excitonlike Polaritons in Semiconductor Microcavities: Effect on the Parametric Scattering of Polaritons

D. N. Krizhanovskii[a], M. N. Makhonin[a], A. I. Tartakovskii[b], and V. D. Kulakovskii[a]

[a]*Institute of Solid State Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*
[b]*Department of Physics and Astronomy, University of Sheffield, Sheffield S3 7RH, UK*
e-mail: kulakovs@issp.ac.ru
Received May 28, 2004

Abstract—Polariton emission in GaAs-based microcavities has been studied under variable conditions, which made it possible to excite (a) polaritons from the upper polariton branch and hot free polaritons and electrons, (b) polaritons from the lower polariton branch (LPB) and localized excitons, and (c) the mixed system. Variation of the excitation conditions leads to substantial differences in the energy distributions of polaritons and in the temperature dependences of polariton emission. It is established that the energy relaxation of resonantly excited LPB polaritons via polariton and localized exciton states at liquid helium temperatures is ineffective. Instead, the relaxation bottleneck effect is suppressed with increasing temperature by means of exciton delocalization (due to thermal excitation by phonons). The most effective mechanism of relaxation to the LPB bottom is via scattering of delocalized excitons on hot free carriers. It is found that the slow energy relaxation of polaritons excited below the free exciton energy can be significantly accelerated at low temperatures by means of additional weak generation of hot excitons and, especially, hot electrons. This acceleration of the energy relaxation of polaritons by means of additional overbarrier photoexcitation sharply decreases the barrier for stimulated parametric scattering of polaritons excited at an LPB inflection point. Therefore, additional illumination can be used to control the polariton–polariton scattering. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Exciton–photon interaction in semiconductor microcavities (MCs) with quantum wells in the active layer leads to the formation of quasi-two-dimensional (2D) exciton–photon states called MC polaritons [1]. These coupled states between photon and exciton are characterized by extremely low density of states (on the order of 10^{-4} of the level for exciton states) and by finite energies at a zero quasi-momentum k. These unique features of the MC polariton dispersion, on the one hand, and short polariton lifetimes, on the other hand, lead to a significant difference of the energy relaxation of polaritons in MC structures from the relaxation of bulk (3D) polaritons and that of 2D excitons in quantum wells.

In particular, strong dependence of the MC polariton energy on the quasi-momentum at small k values significant decreases (similarly to the case of 3D polaritons in semiconductors) the rate of energy relaxation with acoustic photon emission. At a low density of excitation, this results in a strong decrease in the population of states on approaching the bottom of the lower polariton branch (LPB). This phenomenon is known as the bottleneck effect [2]. Calculations showed that, in MCs with a Rabi splitting of $\Omega = 6$ meV and a negative detuning of $\Delta = E_C(k = 0) - E_X \sim -\Omega$ between the photon and exciton mode energies, the population of states at the LPB bottom decreases by three orders of magnitude as compared to the population of states near the free exciton energy E_X [2]. The bottleneck effect was experimentally studied in MCs under conditions of continuous-wave (cw) and pulsed laser excitation [3, 4]. The results of these investigations showed that even weak resonant excitation of polaritons led to a significantly less pronounced decrease in the population of MC polariton states as compared to the calculated values. For example, in MCs with $\Delta \sim -\Omega$ for $\Omega = 6$ meV, the population of polariton states at the LPB bottom decreased only by one order in magnitude [3, 4], which is two to three orders lower than predicted by the theory [2]. This behavior indicates that a considerable role in polariton relaxation is played, besides the scattering on acoustic phonons, by some other scattering mechanisms. These include, first, the scattering due to interparticle interactions. In particular, Malpuech *et al.* [5] found that, in the presence of even a small 2D electron gas density in a quantum well, the scattering of polaritons on electrons leads to a stronger energy relaxation of polaritons than does the scattering on phonons. These calculations qualitatively explained the experimentally observed strong dependence of the polariton energy distribution in the LPB on the conditions of photoexcitation [6, 7].

This paper presents the results of a thorough investigation of the energy relaxation in MCs with InGaAs

quantum wells in the active layer in various regimes of optical excitation, which led to different initial compositions of the photoexcited exciton–electron–polariton system. This was achieved by using lasers generating light quanta with various energies $\hbar\omega_{exc}$: a cw He–Ne laser and two tunable solid-state titanium sapphire (Ti–Sp) lasers. The He–Ne laser (with $\hbar\omega_{exc} \gg E_g$, where E_g is the bandgap width of GaAs) was used to excite hot free carriers (electrons and holes) and excitons. Using tunable Ti–Sp lasers, we produced resonant excitation of polaritons with various energies corresponding to lower and upper polariton branches. Different compositions of the photoexcited exciton–electron–polariton system were obtained by varying the laser output powers.

The results of investigation of the photoluminescence (PL) of polaritons showed that their energy distribution in fact strongly depends on the photoexcitation conditions and the temperature. It was also confirmed that even weak He–Ne laser irradiation during the Ti–Sp-laser-induced resonant excitation of MCs near the LPB inflection point strongly accelerates polariton relaxation to the LPB bottom, which may lead to almost tenfold decrease in a threshold for the stimulated parametric scattering of polaritons.

The paper is organized as follows. In Section 2 we describe samples and experimental techniques. Section 3 reports on the temperature dependence of polariton emission upon resonant excitation to the LPB of the system. Then we will consider how additional resonant excitation to the upper polariton branch (UPB) of the system (Section 4) and additional excitation with an energy exceeding the bandgap in the barrier layer (Section 5) affect the polariton emission upon resonant excitation to the LPB of the system. Section 6 is devoted to the effect of additional nonresonant excitation on the threshold of stimulated parametric scattering of polaritons.

2. SAMPLES AND EXPERIMENTAL TECHNIQUES

We have studied emission from a semiconductor MC grown by molecular-beam epitaxy. The experiments were performed on distributed Bragg reflectors comprising 20 and 17 stacked $Al_{0.13}Ga_{0.87}As/AlAs$ pairs beneath and above the active element, respectively. Each stack consisted of ($\lambda/4$)-thick layers, where λ is the wavelength to which the MC is tuned. The active element comprised six $In_{0.06}Ga_{0.94}As$ quantum wells with a thickness of 70 Å situated in the antinodes of an electromagnetic field in a ($3\lambda/2$)-thick GaAs layer confined between two Bragg mirrors. The GaAs layer had a wedge shape, such that the Fabry–Perot mode energy monotonically varied along the sample at a rate on the order of 0.7 meV/mm. This sample configuration allowed the energy distance between the Fabry–Perot mode and the exciton energy to be varied by exciting luminescence in different points of the sample.

The structure was placed in an optical cryostat into helium vapor. The sample temperature could be controlled in a range from 5 to 30 K to within 0.2 K. Tunable (Ti–Sp) and cw (He–Ne) lasers ($E_{He–Ne} \approx 1.96$ eV) were used for the resonant and nonresonant excitation of polariton emission, respectively. The resonant excitation was performed at a certain angle Φ_{exc} relative to the normal to the MC plane, which provides the excitation of polaritons with $k = q\sin\Phi_{exc}$, where q is the photon wavevector in vacuum. The PL signals were measured using a system comprising a double monochromator (RAMANOR U-1000) and a CCD camera. The measurements were performed at several points of a sample with a Rabi splitting of $\hbar\Omega \approx 6$ meV and a detuning between the photon and exciton mode energies Δ from –2 to –6 meV.

3. EFFECT OF TEMPERATURE ON THE POLARITON EMISSION UPON RESONANT EXCITATION TO THE LOWER POLARITON BRANCH

Figure 1a shows the PL spectra of polariton states recorded at $T = 6$ K (solid curve) and $T = 20$ K (dashed curve). The signal was detected in the direction of the normal to the sample surface, which corresponded to the emission of polaritons with $k = 0$. The resonant excitation of samples to the LPB was performed using laser radiation with a power density of $W_A = 240$ W/cm^2 incident at $\Phi_{exc} = 19°$ ($k = 2.5 \times 10^4$ cm^{-1}). The energy of exciting quanta was $\hbar\omega_{exc} = 1.4519$ eV, that is, 2.6 meV below the exciton energy level ($E_X \approx 1.4545$ eV). As expected, the PL spectrum measured at a lower temperature ($T = 6$ K) displays a single lower polariton (LP) band corresponding to emission from the LPB bottom. Fine structure of the LP band is due to interference within the sample thickness (GaAs substrate is transparent in this spectral range). As the temperature increases to 20 K, the LP band intensity shows a more than tenfold growth, which is evidence of a significant increase in the polariton energy relaxation rate. In addition, the PL spectrum measured at 20 K exhibits an upper polariton (UP) band corresponding to emission from the UPB bottom, which appears due to the scattering of polaritons with phonon absorption.

Figure 1b displays the PL spectra observed for the MC nonresonantly excited by He–Ne laser radiation at a power density of $W_{He–Ne} = 60$ W/cm^2. As can be seen, the LP and UP bands are observed in this case at both 6 and 20 K, but the UP band intensity is still significantly lower than that of the LP band. This ratio of intensities is evidence of a sufficiently effective relaxation of the photoexcited polaritons from the upper to lower branch. For the excitation above E_g, in contrast to the resonant excitation below E_X, the LP band intensity exhibits a decrease (approximately twofold over the 6–20 K inter-

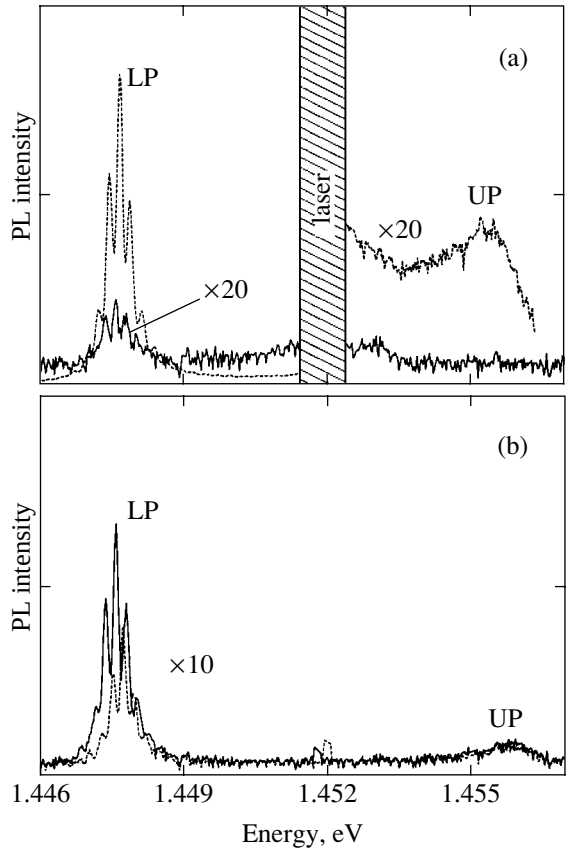


Fig. 1. MC polariton PL spectra measured for $k = 0$ at $T = 6$ K (solid curves) and $T = 20$ K (dashed curves) using (a) resonant excitation into LPB with Ti–Sp laser A ($W_A = 260$ W/cm^2) at $\Phi_{exc} = 19°$ and (b) nonresonant overbarrier excitation above E_g of GaAs with a He–Ne laser ($W_{He–Ne} = 60$ W/cm^2). The LP and UP bands correspond to emission from the LPB and UPB bottom, respectively.

val) with increasing temperature, rather than an increase. The UP band intensity remains virtually unchanged.

The difference in behavior of the LP and UP band intensities in response to the temperature variation for excitation below E_X and E_g is related to the fact that the process of energy relaxation in the MC depends on the composition of the photoexcited exciton–electron–polariton system. Indeed, nonresonant excitation of the MC predominantly leads to the formation of electrons, holes, and excitons in the GaAs barrier, which are trapped by the quantum well to form a system of nonradiative "hot" excitons with large quasi-momenta. These excitons are scattered into polariton states with small k [2], which is followed either by recombination or by trapping on quantum well potential fluctuations with the formation of long-lived (~100 ps) localized excitons [8]. In the case of the resonant excitation to LPB by the Ti–Sp laser radiation with $\hbar\omega_{exc} < E_X$, the excited polaritons are scattered either to the LPB states with smaller k or to the degenerate (with LPB) localized exciton states. In our experiments, the energy of excitation by the Ti–Sp laser A was approximately 2.6 meV below E_X. For this reason, the scattering of resonantly excited polaritons into high-energy exciton states with large k due to the absorption of acoustic phonons at $T = 6$ K ($kT \approx 0.5$ meV) is unlikely.

Thus, the difference in the distribution of polariton population of the LPB states in the region of polariton bottleneck ($k \sim 10^4$ cm^{-1}) observed under variable excitation conditions is related to a difference in the composition of the photoexcited exciton–electron–polariton system. In order to determine the distributions of polaritons with respect to the quasi-momentum under different excitation conditions, we have measured the angular dependences of the LP intensity ($k \sim \sin\Phi$). As can be seen from $I_{LP}(k)$ curves presented in Fig. 2a, the population of LPB states under all excitation conditions decreases on approaching the bottom of the band. This nonequilibrium distribution of polariton states is a consequence of the bottleneck effect cased by a strong dispersion of polaritons and their short lifetimes (comparable to the characteristic time of scattering on acoustic phonons [2–4]). However, the behavior of $I_{LP}(k)$ for the nonresonant excitation above the bandgap is quantitatively different from that for the resonant excitation into LPB below E_X. For the sake of more convenient comparison of the $I_{LP}(k)$ curves at $T = 6$ K in Fig. 2a, the excitation densities of Ti–Sp (260 W/cm^2) and He–Ne (60 W/cm^2) lasers were selected so as to ensure that the LP band intensities would be approximately equal at $k \sim 2 \times 10^4$ cm^{-1}. Under these conditions, the $I_{LP}(k \approx 0)$ value for the nonresonant (He–Ne laser) excitation is about ten times that for the resonant excitation (Ti–Sp laser A). A comparison to the results of calculations [2] showed that a decrease in the population of MC polariton states observed in the experiment using He–Ne laser radiation was three orders of magnitude lower, and that observed for Ti–Sp laser was two orders of magnitude lower than the value predicted by the theory for the phonon mechanism of polariton relaxation.

Now let us consider the effect of temperature on the process of energy relaxation in the polariton system. As can be seen from Fig. 2a, the $I_{LP}(k)$ dependence for nonresonant excitation weakly varies with the temperature. In the case of resonant excitation below E_X, an increase in the temperature from 6 to 20 K is accompanied by significant changes in the population of polariton states near the LPB bottom: the $I_{LP}(k)$ curve becomes less steep and the intensity of emission from LPB for $k = 0$ exhibits a tenfold increase (Figs. 1 and 2). In order to explain the effect of temperature on the polariton emission, let us consider Fig. 2b showing the temperature dependence of $I_{LP}(k = 0)$ for the excitation with Ti–Sp laser radiation. As can be seen, the PL intensity increases with the temperature according to an exponential law, in proportion to $\exp(-\Delta E/kT)$ with an activation energy of $\Delta E \approx 2.6$ eV. This value coincides with

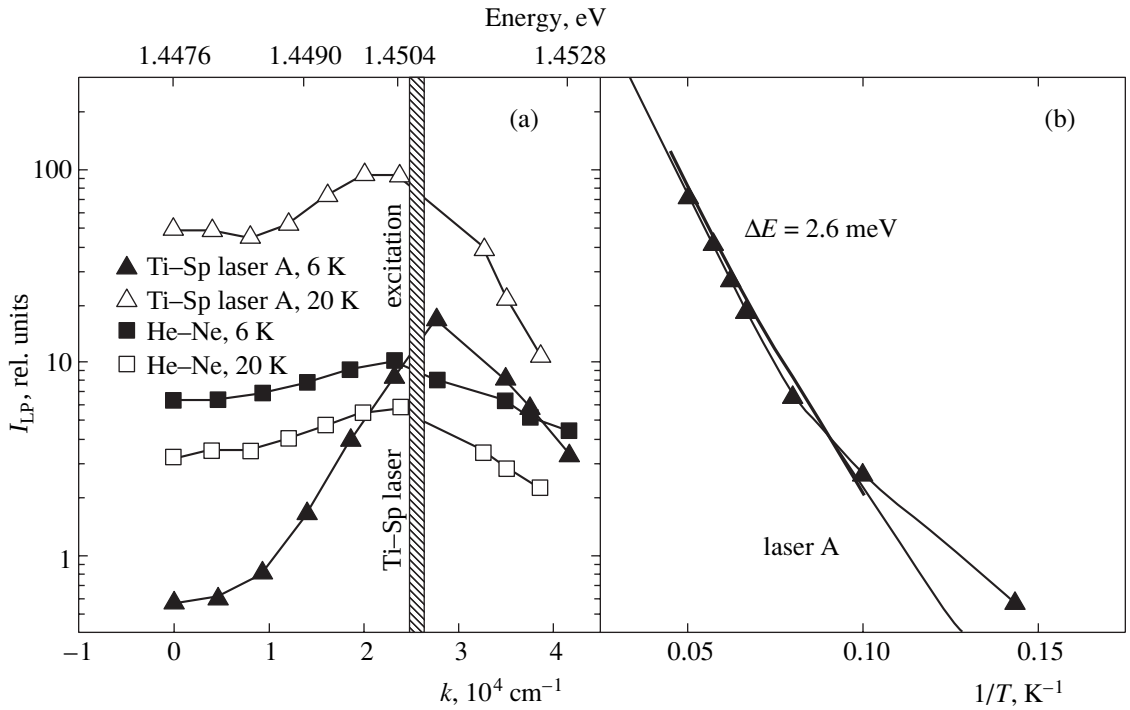


Fig. 2. MC polariton emission: (a) the LP band intensity versus k for the excitation with Ti–Sp laser A ($\hbar\omega_{exc} = E_X - 2.6$ meV) and He–Ne laser radiation; (b) temperature dependence of I_{LP} ($k = 0$) for the excitation with Ti–Sp laser A.

the difference between the exciton level and the excitation energy $E_X - \hbar\omega_{exc} \approx 2.6$ meV. In order to trace variations of the activation energy as dependent on this difference, we measured the temperature dependences of $I_{LP}(k = 0)$ at various Φ_{exc}, which allowed the energy of resonant excitation by the Ti–Sp laser to be varied within $E_X - \hbar\omega_{exc} = 2.1$–$3.1$ meV. As expected, the activation energy was dependent on this energy difference, but the relation $\Delta E \approx E_X - \hbar\omega_{exc}$ was retained. Thus, we may suggest that an additional channel of polariton relaxation at higher temperatures is related to the thermal activation of long-lived (~100 ps [8]) localized exciton states photoexcited upon the resonant excitation into the LPB at $E \approx \hbar\omega_{exc}$. A strong decrease in the intensity of polariton emission observed on approaching the LPB bottom at a lower temperature indicates that the relaxation of such localized excitons with the emission of acoustic phonons, both into low-energy LPB states and into low-energy states of localized excitons followed by the scattering into LPB states, is not an effective mechanism.

The characteristic exponential behavior of the polariton emission intensity as a function of the temperature indicates that the relaxation of resonantly excited polaritons and localized excitons at 20 K proceeds in two stages. In the first stage, thermal activation of the localized excitons leads to occupation of the low-energy excitonlike states in the LPB. At the second stage, these excitons are scattered into states at the LPB bottom. Note that the absorption of Ti–Sp laser radiation in the range of temperatures $T < 20$ K remains almost constant (relative variation $\Delta A/A$ does not exceed 20%). Therefore, a contribution to the optical absorption (upon the resonant excitation) due to transitions into the LPB with the absorption of acoustic phonons is relatively small. Thus, the optical absorption coefficient is predominantly determined by the scattering into localized excitons, while an increase in the polariton emission intensity I_{LP} with the temperature is directly related to delocalization of the photoexcited localized excitons.

4. EFFECT OF ADDITIONAL RESONANT EXCITATION TO THE UPB ON THE PHOTOLUMINESCENCE OF POLARITONS RESONANTLY EXCITED TO THE LPB

In order to elucidate mechanisms responsible for the energy relaxation of excitonlike polaritons, we have studied the PL of polaritons under conditions of simultaneous excitation into UPB and LPB, whereby localized excitons and polaritons in the LPB, polaritons in the UPB, and hot free excitons with large quasimomenta were excited. The ratio between various excited states could be changed by varying the relative density of excitation into the LPB and UPB.

Figure 3a shows the PL spectra of MCs with a detuning between the exciton (X) and photon (C) modes $\Delta = E_X - E_C \approx -5.3$ meV. The emission was

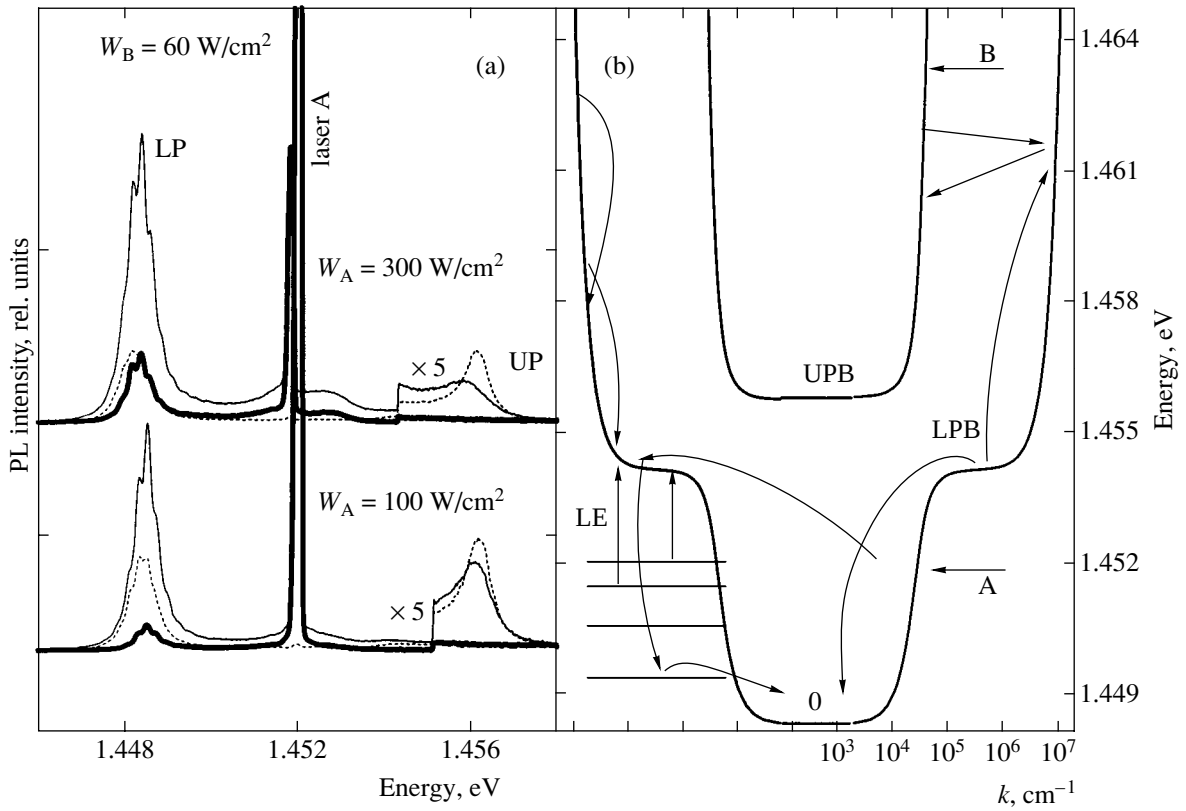


Fig. 3. (a) MC polariton emission spectra measured for $k = 0$ at $T = 6$ K using resonant excitation into UPB with Ti–Sp laser B at $\Phi_B = 28°$, $W_B = 60$ W/cm^2 (dashed curves), into LPB with Ti–Sp laser A at $\Phi_A = 19°$, $W_A = 100$ or 300 W/cm^2 (thick solid curves), and with both Ti–Sp lasers (thin solid curves); (b) schematic diagram of energy relaxation in the exciton–polariton system upon simultaneous resonant excitation with lasers A and B into the LPB and UPB, respectively.

detected along the normal to the MC plane. The measurements were performed at $T = 6$ K under various excitation conditions. Thick solid curves represent the spectra observed for a resonant excitation into the LPB with the Ti–Sp laser A at an excitation angle of $\Phi_A = 19°$ ($\hbar\omega_A = 1.4519$ eV $\approx E_X - 2.6$ meV). As was pointed out above (Fig. 1), the single predominant signal in these spectra is due to emission from the LPB bottom. Dashed curves represent the PL spectra recorded using a resonant excitation into the UPB with the Ti–Sp laser B at an excitation angle of $\Phi_B = 28°$ ($\hbar\omega_B = 1.4633$ eV). These spectra display the signals due to both lower (LP) and upper (UP) polaritons. Thin solid curves in Fig. 3a show the spectra of emission along the normal to the MC plane measured using simultaneous excitation of the sample with Ti–Sp lasers A and B ($\hbar\omega = 1.4519$ and 1.4633 eV, respectively). The two families of spectra presented in Fig. 3a were recorded for two different densities of excitation using laser A ($W_A = 100$ and 300 W/cm^2, respectively) at constant densities of excitation using laser B ($W_B = 60$ W/cm^2). A comparison of the spectra observed using the excitation with laser B only and with both lasers A and B shows that the contribution to the intensity of polariton emission due to the latter laser is not additive: an additional excitation to the LPB leads to an increase in intensity of the LP band and a decrease in intensity of the UP band, that is, stimulating the polariton relaxation from UPB to LPB. In particular, Fig. 3a shows that, for $W_A = 100$ W/cm^2, the intensity of the LP band obtained by excitation with two lasers (I_{LP}^{A+B}) is about 1.5 times the sum of intensities of the LP bands observed upon separate excitation with lasers A and B ($I_{LP}^{A} + I_{LP}^{B}$). As the power of laser A is increased to 300 W/cm^2, the ratio $I_{LP}^{A+B}/(I_{LP}^{A} + I_{LP}^{B})$ increases to 2.5.

Figures 4a and 4b show plots of the PL intensity for polaritons from the LPB and UPB bottom versus density of excitation into the UPB, measured with a fixed density of excitation into the LPB ($W_A = 300$ W/cm^2) and without such excitation, respectively. The dashed line in Fig. 4a shows the intensity of emission from the LPB bottom in the case of excitation using only laser A at $W_A = 300$ W/cm^2. As can be seen from these data, the excitation into the UPB only (open circles) leads to a slightly sublinear increase in the UP band intensity and to a slightly superlinear growth in the LP band intensity. These deviations from linearity indicate that both the

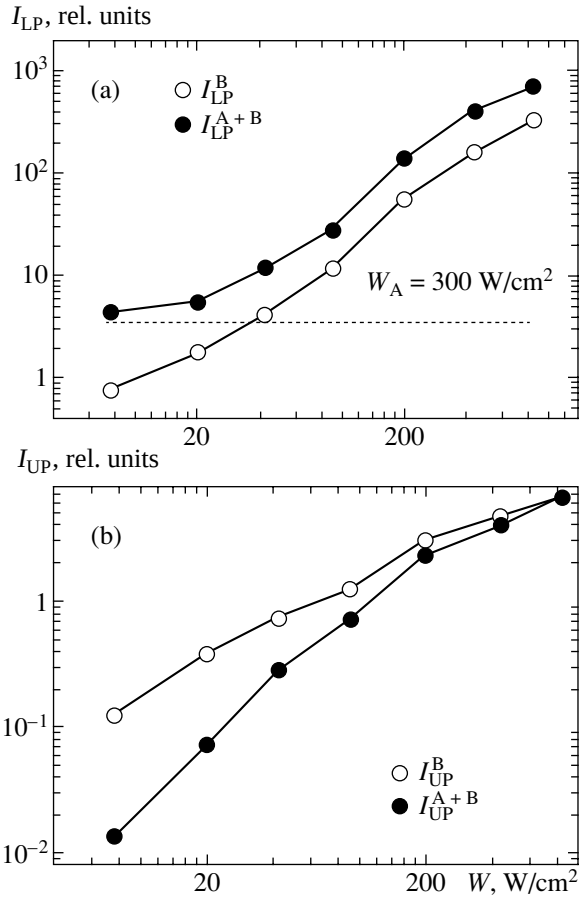


Fig. 4. Plots of the intensity of polariton emission from (a) LPB and (b) UPB bottom versus density of excitation into the UPB, measured without (I_{LP}^{B}, I_{UP}^{B}) and with (I_{LP}^{A+B}, I_{UP}^{A+B}) a fixed density of excitation into the LPB (W_A = 300 W/cm^2). The dashed line in Fig. 4a shows the intensity of emission from the LPB bottom in the case of excitation using only laser A at W_A = 300 W/cm^2.

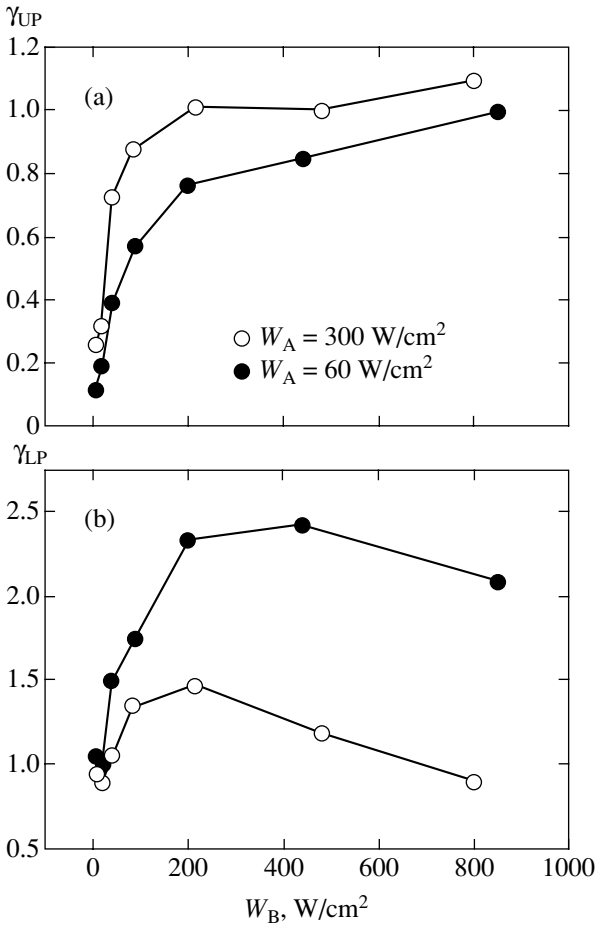


Fig. 5. Plots of the ratio of the PL intensity for the MC simultaneously excited with lasers A and B to the sum of intensities for the separate excitation with each laser versus power density of laser B for the emission from (a) UPB bottom ($\gamma_{UP} = I_{UP}^{A+B}/I_{UP}^{B}$) and (b) LPB bottom ($\gamma_{LP} = I_{LP}^{A+B}/(I_{LP}^{A} + I_{LP}^{B})$) for W_A = 60 and 300 W/cm^2.

rate of the relaxation from the UPB to LPB and the rate of the energy relaxation to the LPB bottom increase with the density of excited polaritons. An increase in the relaxation rate is related to interparticle collisions. These collisions also account for the difference in intensity of the LP and UP bands recorded upon the simultaneous (I_i^{A+B}) and separate ($I_i^{A} + I_i^{B}$) MC excitation using two Ti–Sp lasers. Figure 4 also shows that the difference between $I_{UP}^{A} + I_{UP}^{B}$ and I_{UP}^{A+B} for the emission from the UPB is maximum under the conditions of a small density of excitation into the UPB. As for the LP band intensity, the difference between $I_{LP}^{A} + I_{LP}^{B}$ and I_{LP}^{A+B} exhibits the opposite trend and increases with the density of excitation into the UPB. This behavior is quantitatively characterized in Fig. 5, which shows plots of the ratios $\gamma_{LP} = I_{LP}^{A+B}/(I_{LP}^{A} + I_{LP}^{B})$ and $\gamma_{UP} = I_{UP}^{A+B}/I_{UP}^{B}$ versus the power of laser B for W_A = 60 and 300 W/cm^2. As can be seen, the maximum value of γ_{LP} (~1.5) for W_A = 60 W/cm^2 is achieved at W_B ~ 240 W/cm^2, while for W_A = 300 W/cm^2, the maximum (~2.5) is achieved at W_B ~ 500 W/cm^2.

Figure 6a presents a plot of the intensity of emission from the UPB bottom versus quasi-momentum k measured at T = 6 K for the resonant excitation into the UPB with laser B operating at W_B = 60 W/cm^2. As can be seen, the energy distribution of polaritons over the UPB (in contrast to that over the LPB) is close to thermodynamically equilibrium distribution for a certain effective temperature: the density of polaritons exponentially increases on approaching the UPB bottom even at very low excitation densities. Apparently, such a distribution cannot be established due to the relaxation of

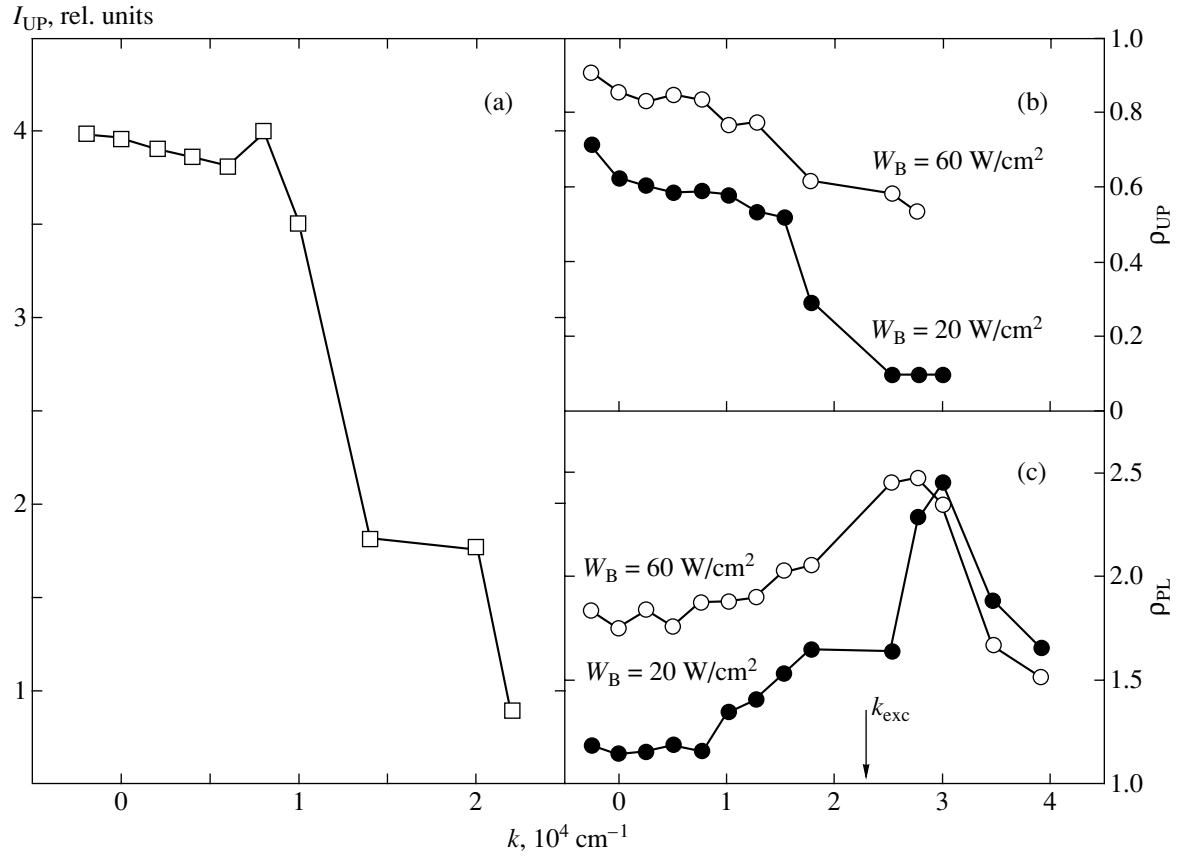


Fig. 6. Plots of (a) the intensity of polariton emission from UPB versus quasi-momentum k measured at T = 6 K upon resonant excitation into UPB with laser B (W_B = 60 W/cm^2) and (b, c) the relative intensity of emission from the UPB ($\rho_{UP} = I_{UP}^{A+B}/I_{UP}^{B}$) and LPB ($\rho_{LP} = I_{LP}^{A+B}/(I_{LP}^{A} + I_{LP}^{B})$), respectively, versus quasi-momentum k for W_B = 20 and 60 W/cm^2.

polaritons with the emission of acoustic phonons only for the UPB. Indeed, on the one hand, the UPB dispersion is not less steep than the LPB dispersion and, hence, the relaxation to the UPB bottom, as well as to the LPB bottom, requires the emission of several phonons. On the other hand, the polariton lifetime on the UPB is (because of a significant photon contribution) even shorter than that on the LPB. A quasi-equilibrium distribution of polaritons on the UPB can be explained only by including into consideration the processes of scattering between UPB and LPB. Indeed, the scheme in Fig. 3b shows that the most probable process even at the first stage is the scattering of polaritons (excited into the excitonlike states with large quasi-momenta on the UPB) to the LPB with the emission or absorption of acoustic phonons [2]. First, the density of exciton states is several orders of magnitude greater than the density of states on the UPB; second, the probability of polariton scattering with the emission of phonons is determined by the relative fraction of the exciton component, which is small for the UPB states and close to unity for the high-energy LPB states. The exciton states on the LPB with $k > k_0 \sim 10^5$ cm^{-1} are optically inactive and characterized by very large (nanoseconds) lifetimes [9]. Since these lifetimes are greater by several orders of magnitude than the characteristic times of the exciton–exciton collisions and the scattering on acoustic phonons, the exciton system acquires a distribution with a certain effective temperature close to the lattice temperature [10, 11]. The energy distribution of polaritons at the UPB bottom is formed due to the reverse scattering of equilibrium long-lived polaritons from LPP to UPB. Since the energies of acoustic phonons involved in the scattering process fall within 1 meV [2], the distribution of polaritons in the UPB measured in experiment reflects with sufficient accuracy the process of energy relaxation and the effective temperature of the excitonlike polaritons in the LPB. Thus, by measuring the energy distribution of polaritons in the UPB, it is possible to obtain direct information about the behavior of "dark" excitonlike polaritons in the LPB and to trace variations in the system excited in various regimes.

Now let us consider variations in the distribution of polaritons over the LPB and UPB under the conditions of resonant excitation with two lasers. Figure 6b shows

a change in the distribution of the relative intensity of emission from the UPB, expressed as $\rho_{UP} = I_{UP}^{A+B}/I_{UP}^{B}$, with increasing quasi-momentum, which directly reflects the change in population of the corresponding states. As can be seen, an additional excitation into the LPB below E_X produces two effects: first, a decrease in the integral intensity of emission from the UPB (and, hence, in the total density of polaritons on the UPB) and, second, a more pronounced decrease in intensity with increasing k (or the energy), reflecting a decrease in the effective temperature of polaritons. More effective cooling was observed for a lower density of excitation with laser B, that is, for a lower density of hot excitons and UPB polaritons injected into the system. Taking into account the above considerations concerning relations between the energy distribution of particles between the UPB and LPB, we may conclude that the injection of polaritons with energies below E_X also leads to a decrease in the effective temperature in the exciton system initially excited above E_X. The energy of hot excitons is spent for exciting low-energy polaritons on the LPB and localized excitons. This must result in increasing density of excitonlike polaritons on the LPB in the region of energies $E \sim E_X$ with $k \sim 3 \times 10^4$–10^6 cm^{-1}.

Figure 6c shows the experimentally measured changes in distributions of the LPB population expressed as $\rho_{LP} = I_{LP}^{A+B}/(I_{LP}^{B} + I_{LP}^{A})$. As can be seen, the maximum increase in the polariton density caused by the second laser excitation is observed exactly in the region of $k \sim (2–3) \times 10^4$ cm^{-1}, that is, for energies slightly below the exciton energy level. Moreover, at a low density of excitation into the UPB ($W_B = 20$ W/cm^2), an increase in I_{LP}^{A+B} relative to the sum $I_{LP}^{A} + I_{LP}^{B}$ is observed only in this region of quasi-momenta; only for a greater excitation density ($W_B = 60$ W/cm^2) does this increase spread toward lower quasi-momenta near the LPB bottom.

An analysis of changes in the energy distribution of polaritons caused by variations of the density of excitation into the UPB and LPB showed that the interaction between the two photoexcited systems leads to a nonadditive increase in the LPB population near $E = E_X$ as a result of the cooling of hot excitons produced by laser B and the delocalization of (localized) excitons generated by laser A. The mechanisms of this delocalization include the scattering on phonons (on increasing the temperature) and on hot excitons and free carriers (on illumination with laser B). The most effective delocalization of the localized excitons is caused by their interaction with electrons, since the matrix element of the exciton–electron interaction is several orders of magnitude greater than those of the exciton–exciton and exciton–phonon interactions [5]. Indeed, intentionally undoped MBE-grown semiconductor microstructures are characterized by the density of residual impurity centers not less than 10^{13} cm^{-3}, which leads to the appearance of a 2D electron gas with a density of 10^9–10^{10} cm^{-2} in quantum wells. An increase in the temperature or laser-induced generation of hot excitons leads to heating of the equilibrium electrons and a change in their density due to the deionization of charged centers by the hot phonons and excitons. The scattering of hot free carriers on resonantly excited localized excitons and polaritons leads to an effective population of the LPB near $E = E_X$. The delocalization of excitons requires an energy of $\delta E > E_X - \hbar\omega_{exc} \sim 2.6$ meV.

Thus, the effect produced by additional resonant MC excitation into the UPB is analogous to that of the temperature for the resonant excitation into the LPB, leading to an increase in the density of mobile excitons in the LPB near $E = E_X$. These excitons are effectively scattered to the LPB bottom as a result of exciton–exciton and exciton–electron collisions. The predominant mechanism is the exciton–electron scattering, which is characterized by a large cross section; in addition, the energy and momentum conservation laws impose much lower limitations on the exciton–electron interaction with respect to the energy of excitons which can be effectively scattered to the LPB bottom. For the exciton–exciton scattering, the threshold condition for the kinetic energy of excitons scattered to the LPB bottom can be written as $2E_1E_2 > (E_X - E_{LP}(k))^2/4$, where $E_{1,2}$ are the energies of excitons on the LPB measured from E_X. This formula shows that, in an MC with an LPB depth of $E_X - E_{LP}(k = 0) \approx 5$ meV, effective scattering to the LPB bottom will take place only for particles with energies above 1.7 meV. In the exciton–electron system with the same temperatures of excitons and electrons, effective exciton–electron scattering to the LPB bottom begins with an exciton energy of $E \sim (E_X - E_{LP}(k))m_e/m_X$, where m_e and m_X are the effective masses of electron and exciton, respectively. Since the m_e in GaAs is almost ten times as small as m_X, the scattering to the LPB bottom becomes possible for particles with a much lower threshold energy (on the order of 0.3–0.4 meV) than that for the exciton–exciton interaction.

In order to assess how much the efficiency of scattering to the LPB bottom depends on the kinetic energy of polaritons initially excited to the UPB, we studied the intensity of the polariton emission from the LPB bottom as a function of $\hbar\omega_B$. In these experiments, we changed both the energy $\hbar\omega_B$ and the angle of excitation for laser B in order to realize the conditions of resonant photoexcitation into the UPB. The results of these measurements are plotted in Fig. 7 as the intensities of polariton emission from the LPB at $k = 0$ versus $\hbar\omega_B$, measured for the excitation with both lasers A and B (I_{LP}^{A+B}) and with laser B alone (I_{LP}^{B}). As can be seen, additional excitation to the UPB bottom ($\hbar\omega_B - E_X \sim$ 0.36 meV) leads to addition of intensities observed for

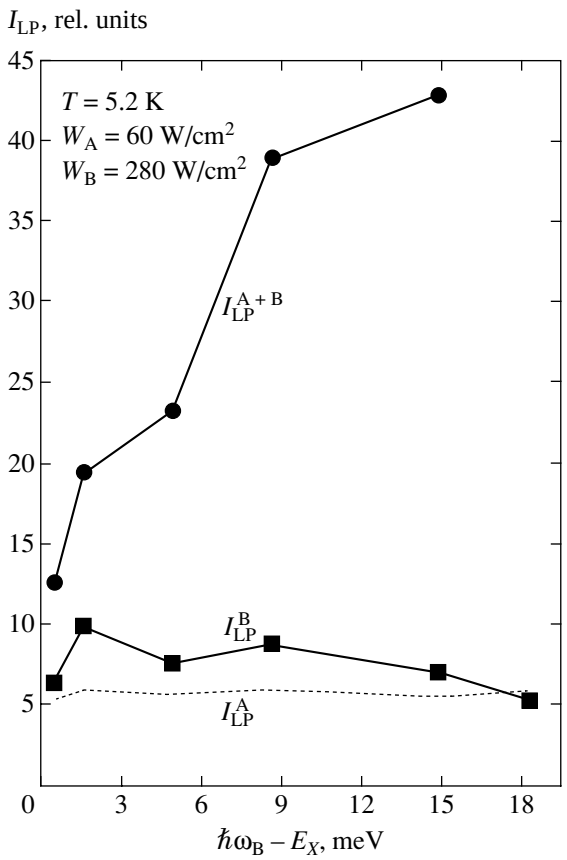


Fig. 7. Plots of the intensity of polariton emission from the LPB at $k = 0$ versus $\hbar\omega_B$ for an MC excited with both lasers A and B (I_{LP}^{A+B}) and with laser B alone (I_{LP}^{B}). Dashed curve shows the intensity I_{LP}^{A} observed for the excitation using only laser A.

the separate excitation using lasers A and B: $I_{LP}^{A+B} = I_{LP}^{A} + I_{LP}^{B}$. However, as the $\hbar\omega_B$ value increases, the ratio of the two intensities grows to reach about three for $\hbar\omega_B - E_X \approx 15$ meV.

An increase in $\hbar\omega_B$ leads primarily to a growth in the average energy of excitons produced by laser B and to an additional heating of the electron gas. In turn, the greater the quasi-momenta of excitons (k_X) and electrons (k_e), the more effective the exciton–electron scattering to the LPB bottom (however, this increase is relatively small). In particular, Fig. 7 shows that, for excitation into the UPB alone, the intensity of polariton emission from the LPB bottom is almost independent of $\hbar\omega_B - E_X$. For this reason, we believe that the lack of additivity in I_{LP} observed with increasing $\hbar\omega_B$ is related to the involvement of localized excitons produced by laser A into the scattering process. The scattering of hot excitons and electrons, excited by laser B, on the localized excitons leads to their delocalization. The estimates presented above for an MC with an LPB depth of about 5 meV show that effective scattering to the LPB bottom is possible for excitons with $k \gtrsim 5 \times 10^6$ cm^{-1}. Upon excitation by laser B to the UPB bottom, the energy of photoexcited excitons is insufficient for the delocalization of excitons produced by laser A (2.6 meV below E_X). For this reason, the scattering of particles, excited by both lasers, to the LPB bottom proceeds independently and the corresponding intensities are additive. As $\hbar\omega_B$ increases, the energy of particles excited by laser B becomes sufficient for the delocalization of excitons excited by laser A. Since the efficiency of the exciton–electron scattering for delocalized excitons is several orders of magnitude higher than that for localized ones, the delocalization of excitons produced by laser A immediately gives rise to density of polaritons on the LPB bottom.

5. EFFECT OF NONRESONANT ABOVE THE BANDGAP LASER EXCITATION ON THE ENERGY RELAXATION OF POLARITONS

It was demonstrated in the previous section that the effect of hot excitons excited by laser B on the relaxation rate of polaritons excited by laser A increases with the kinetic energy of excitons (Fig. 7). It can be expected that the linear dependence of I_{LP} on $\hbar\omega_B$ (Fig. 7) becomes saturated only for $\hbar\omega_B - E_X > \hbar\Omega_{LO}$, that is, when the photoexcited particles will effectively relax with the emission of optical phonons $\hbar\Omega_{LO} \sim 35$ meV. Thus, the maximum effect of enhancement of the energy relaxation of polaritons resonantly excited into the LPB should be expected in the case of an additional overbarrier excitation (with an energy above the GaAs bandgap). Under these conditions, the system will feature the excitation of both hot excitons and free carriers (via the production of electron–hole pairs and the delocalization of impurity centers). The excitation above E_g(GaAs) was provided by He–Ne laser radiation.

Figure 8 shows the behavior of the intensity of polariton emission from the LPB bottom under conditions of simultaneous excitation using two lasers: (i) resonant excitation below E_X into the LPB with a Ti–Sp laser and (ii) nonresonant excitation above E_g(GaAs) with a He–Ne laser. The experiments were performed with variation of the energy of either the Ti–Sp laser (Fig. 8b) or the He–Ne laser (Fig. 8a). As can be seen from Fig. 8b, a relatively weak additional excitation above the bandgap can provide for a more than tenfold increase in the density of polaritons on the LPB bottom, which is significantly higher than the effect provided by a resonant excitation into the UPB with $\hbar\omega_B - E_X < 15$ meV. Thus, it has been established that a relatively weak additional excitation using a He–Ne laser leads to a sharp increase in the rate of polariton energy relaxation to the LPB bottom. Here, a question naturally arises concerning the optimum ratio of the densities of excitation below E_X and above

E_g(GaAs). The acceleration of relaxation to the LPB bottom is related to delocalization of the (localized) excitons excited by laser A with energy below E_X. To a first approximation, the rate of delocalization of the localized excitons is $\tau_{del}^{-1} \propto n_{hot} \propto W_{He–Ne}$, where n_{hot} is the density of hot free carriers excited by the He–Ne laser. Therefore, the density of localized excitons must grow linearly with $W_{He–Ne}$ until τ_{del} becomes comparable to the characteristic time of radiative recombination of the localized excitons (τ_{LE}). With further increase in $W_{He–Ne}$, the density of polaritons on the LPB bottom must reach saturation. It should be noted that, to a first approximation, the rate of delocalization, τ_{del}^{-1}, must be independent of the density of delocalized excitons and, hence, the fraction of delocalized excitons must be independent of the density of resonant excitation below E_X as long as $\tau_{del}^{-1} < \tau_{LE}^{-1}$).

These predictions agree well with the experimental data presented in Fig. 8. As can be seen, the intensity of polariton emission from the LPB bottom in the selected range of variation of the laser power is virtually a linear function of this power for the separate excitation using both He–Ne and Ti–Sp lasers. This behavior implies that neither polariton–polariton nor polariton–exciton interactions play any significant role in the energy relaxation of polaritons for the separate excitation using laser A or the He–Ne laser. However, simultaneous excitation with both lasers leads to a more complicated behavior of the polariton emission intensity $I_{LP}^{He–Ne + A}$. As can be seen from Fig. 8a, $I_{LP}^{He–Ne + A}$ is greater than the sum $I_{LP}^{He–Ne} + I_{LP}^{A}$ even for $W_{He–Ne} \approx 2$ W/cm^2. As $W_{He–Ne}$ further increases up to 30 W/cm^2, the $I_{LP}^{He–Ne + A}$ value exhibits a superlinear increase with the laser power and then reaches saturation for $W_{He–Ne} >$ 70 W/cm^2. The dependence of $I_{LP}^{He–Ne + A}$ on the power density W_A of the Ti–Sp laser A at a fixed power of the He–Ne laser is virtually linear in the entire range of W_A (Fig. 8b). The dependence of the factor of increase in the polariton emission intensity, defined as $\gamma = (I_{LP}^{A + He–Ne} - I_{LP}^{A})/I_{LP}^{A}$, on $W_{He–Ne}$ and W_A is presented in Fig. 9a and 9b, respectively. As can be seen from these data, γ is virtually independent of W_A; at the same time, $\gamma \propto W_{He–Ne}$ for $W_{He–Ne} < 30$ W/cm^2; and γ reaches saturation for $W_{He–Ne} > 70$ W/cm^2.

The rate of delocalization of the (localized) excitons by free carriers can be evaluated using a simple relation,

$$\tau_{del}^{-1} \approx n_{hot} \sigma v,$$

where $\sigma \approx \pi a_X^2$ is the exciton scattering cross section, $a_X \approx 50$ Å is the Bohr radius of exciton, and $v = \sqrt{2E_{kin}/m_e}$ is the velocity of hot free electrons. In the case of excitation using a He–Ne laser, the density of carriers per quantum well is approximately $n \sim 3 \times 10^7$ cm^{-2} W [W/cm^2]. For E_{kin} = 5–10 meV and $W_{He–Ne}$ = 30 W/cm^2, we obtain $\tau_{del} \approx 0.1$ ns. This value is comparable in order of magnitude with the lifetime of localized excitons (τ = 0.3–0.5 ns) [8].

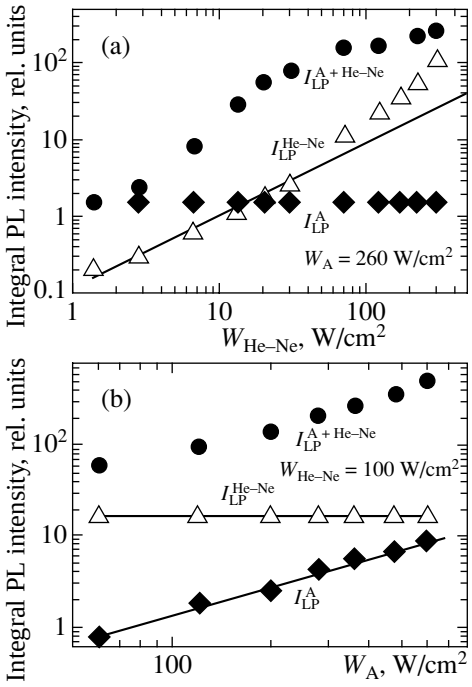


Fig. 8. Plots of the polariton emission intensity from the LPB for $k = 0$ at $T = 6$ K versus (a) $W_{He–Ne}$ and (b) W_A for an MC excited using a He–Ne laser (△), Ti–Sp laser A (◆), and both lasers simultaneously (●).

In addition to relaxation of the delocalized excitons to the LPB bottom as a result of the direct exciton–exciton or exciton–electron scattering from the reservoir of excitons with large k, the LPB bottom can be also populated as a result of the trapping of mobile excitons by deeper localized levels, followed by their scattering (elastic and involving one phonon) to the polariton states (see Fig. 3b). The efficiency of this channel is related to two factors: on the one hand, thermal delocalization of the deep localized excitons is practically absent; on the other hand, the time of scattering of these excitons into the polariton states is comparable to the time of their radiative recombination (0.3–0.5 ns). However, the density of localized states decreases with distance from E_X according to the Gauss law, $A\exp(-(E - E_X)^2/\delta_L^2)$. In the MC structure studied (six quantum wells), the halfwidth of the inhomogeneous exciton band was δ_L = 1.5 meV and the density of localized states became lower than the density of polariton states (on the order of 2×10^7 cm^{-2} meV^{-1}) already at $E = E_X - 5$ meV. Thus, it can be expected that, in the

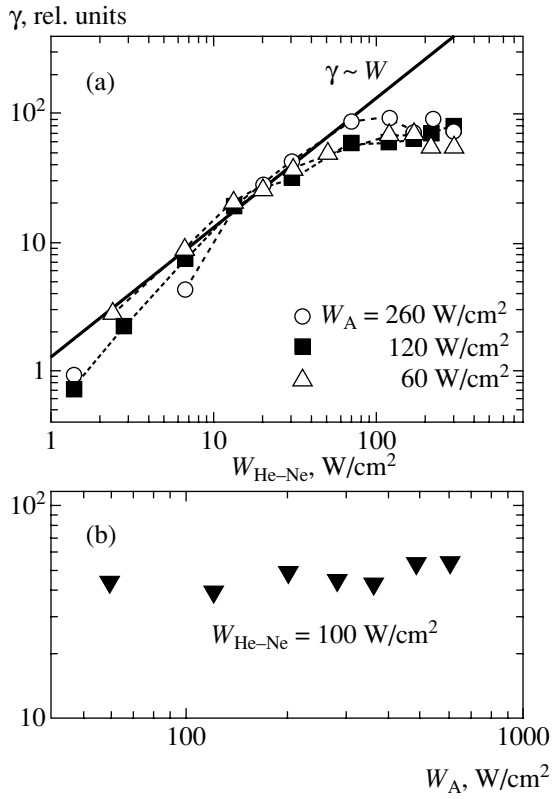


Fig. 9. Plots of $\gamma = (I_{LP}(A + \text{He–Ne}) - I_{LP}(A))/I_{LP}(A)$ versus (a) $W_{He–Ne}$ and (b) W_A; γ is the factor of increase in the emission intensity from the LPB for $k = 0$ at $T = 6$ K, caused by simultaneous excitation using He–Ne and Ti–Sp lasers.

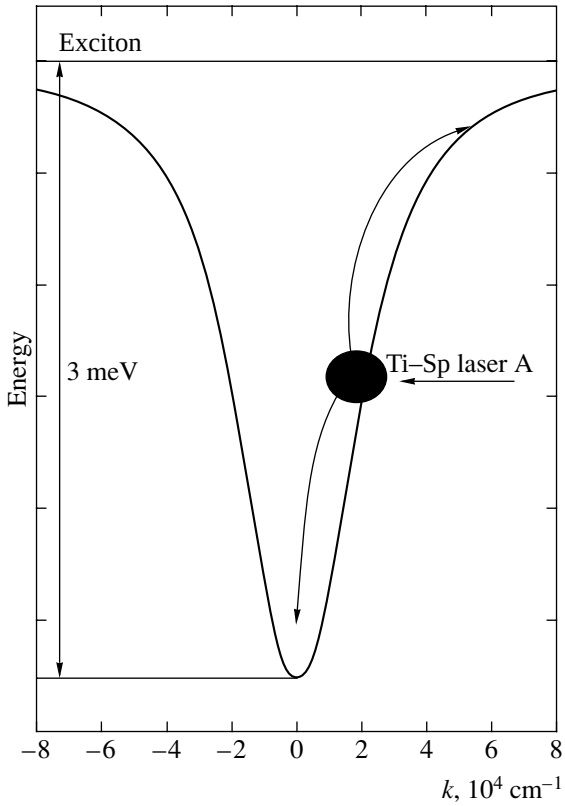


Fig. 10. Schematic diagram of parametric scattering, showing the distance from the LPB bottom to the free exciton energy level.

structures with relatively shallow LPBs, the relaxation into polariton states at the LPB bottom competes with the direct scattering of mobile excitons.

6. PARAMETRIC SCATTERING OF POLARITONS STIMULATED BY MEANS OF NONRESONANT LASER EXCITATION

Earlier [12–15] it was established that a polariton system, in contrast to exciton systems, may exhibit intense parametric scattering under certain excitation conditions for relatively low densities of excitation in MCs with shallow LPBs ($E_X - E_{LP} \approx 2$–4 meV). In particular, excitation into the LPB with k_p near the bending point provides conditions for polariton–polariton scattering into the states near $k = 0$ and $k = 2k_p$ with observation of the laws of energy and quasi-momentum conservation: $2E(k_p) = E(k = 0) + E(2k_p)$. The scheme of this scattering is presented in Fig. 10. When the excitation density increases to a certain threshold level W_{th}, the scattering acquires stimulated character [12–15]. At $T = 5$ K, the efficiency of the stimulated polariton–polariton scattering to the LPB bottom upon excitation near the inflection point even at a moderate power density ($W \sim 100$–200 W/cm^2) exceeds by several orders of magnitude the efficiency of the polariton–phonon scattering considered above [12–15]. In order to provide conditions for the stimulated parametric scattering, it is necessary to ensure large populations of the final polariton states at the LPB bottom. Thus, a question arises as to whether it is possible to modify the threshold of the stimulated scattering of polaritons by means of a weak additional nonresonant photoexcitation. It was demonstrated above that this leads to a significantly accelerated relaxation to the LPB bottom for the polaritons initially scattered to the localized states with $E \approx \hbar\omega_A$. In order to answer this question, we studied the effect of an additional above-bandgap excitation on the emission from the LPB bottom in MCs with $E_X - E_{LP} \approx 4$ meV resonantly excited by a Ti–Sp laser near the inflection point $E_{LP}(k)$.

Figure 11a shows a distribution of the polariton emission intensity with respect to the quasi-momentum for excitation near the inflection point at an excitation density about 1.5 times below the critical density for the development of stimulated parametric scattering. Under these conditions, the intensity of emission exhibits an almost tenfold monotonic decrease on approaching the LPB bottom. Figures 11b and 11c demonstrate a change in the emission intensity distribution under the conditions of a weak additional photoexcitation produced by a He–Ne laser. As can be seen, the pattern is

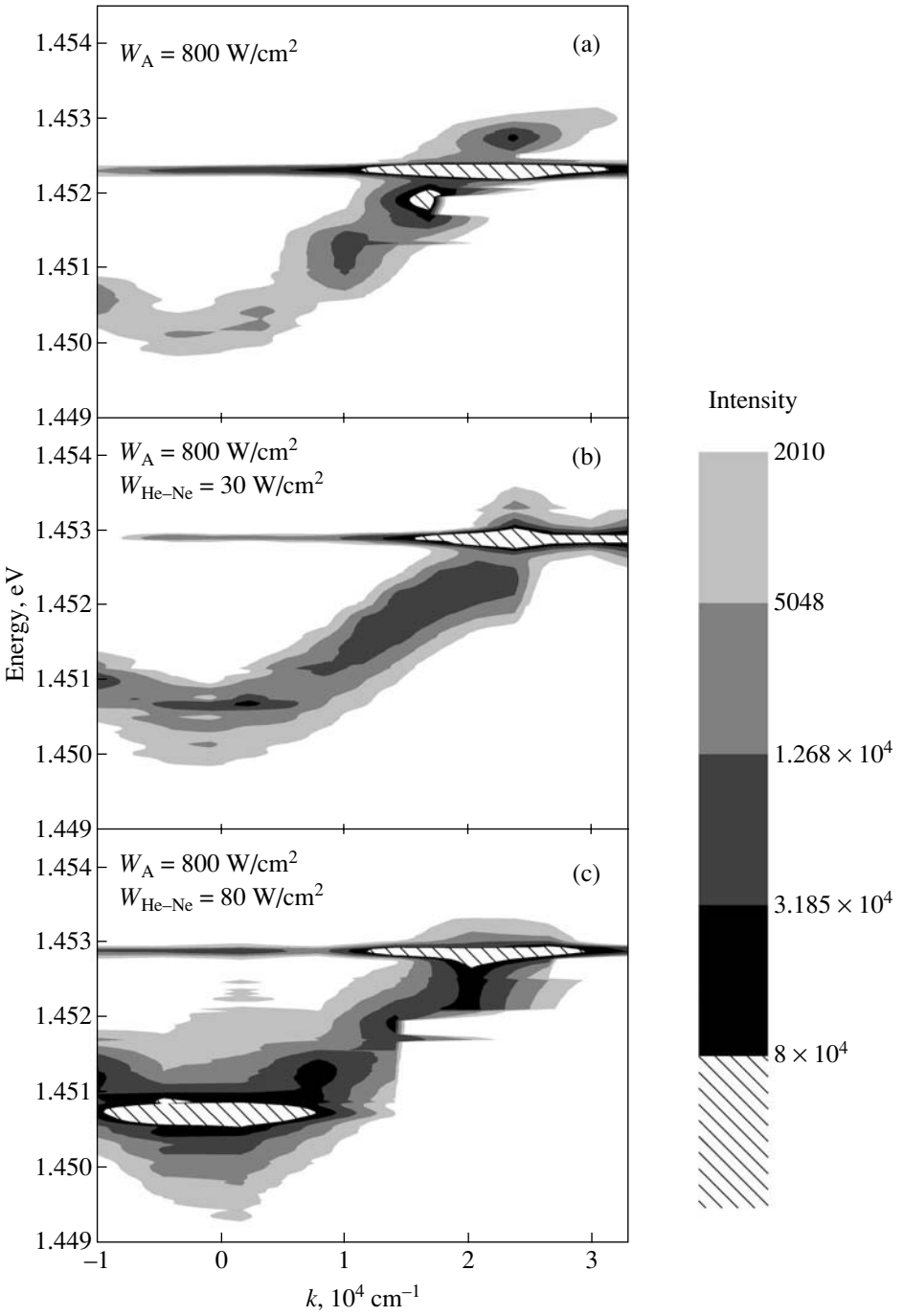


Fig. 11. The intensity of polariton emission from the LPB bottom versus quasi-momentum and energy for the resonant excitation into the LPB near the bending point (W_A = 800 W/cm^2; T = 5.2 K): (a) without additional illumination; (b, c) with additional excitation using a He–Ne laser with $W_{He–Ne}$ = 30 and 80 W/cm^2, respectively.

modified even upon an additional illumination at $W_{He–Ne}$ = 30 W/cm^2, for which the intensity is almost independent of the quasi-momentum. As the laser power density increases to $W_{He–Ne}$ = 80 W/cm^2, the distribution exhibits a strong maximum at $k \approx 0$ with an angular width below 3°, which is evidence of the development of a parametric scattering into states with $k \sim 0$. It should be noted that the intensity of emission in the case of excitation using the He–Ne laser alone is almost ten times lower than that observed for the resonant excitation, so that excitation by the He–Ne laser cannot ensure macroscopic population of the LPB bottom. Therefore, macroscopic population of the LPB bottom is reached due to stimulation of the energy relaxation of

localized excitons (excited below E_X by the Ti–Sp laser) to the LPB bottom as a result of the interaction with photoexcited hot excitons and electrons. Thus, by means of injection of hot electrons at a relatively low density into the polariton system, it is possible to control the threshold for stimulated parametric scattering in the polariton system of planar semiconductor MCs.

7. CONCLUSIONS

Investigations of the relaxation of a polariton system in a GaAs MC photoexcited below the free exciton energy level at $T = 6$ K showed that the relaxation of polaritons over the LPB and over the states of localized excitons with the emission of acoustic phonons is ineffective. An increase in temperature leads to suppression of the bottleneck effect in polariton relaxation. It was established that the energy relaxation to the LPB bottom is a two-stage process. At the first stage, excitons are delocalized as a result of the thermal activation by phonons; at the second stage, these delocalized excitons are scattered into states at the LPB bottom, either as a result of direct exciton–exciton or exciton–electron scattering into polariton states, or due to exciton localization on deeper fluctuations, followed by their scattering into LPB states (in MCs with relatively shallow LPBs).

At low temperatures, the energy relaxation of polaritons excited below the free exciton energy level can be significantly accelerated by means of additional weak generation of hot excitons and, especially, hot electrons. The collisions of hot excitons and electrons with resonantly excited excitons and LPB polaritons lead, on the one hand, to a strong cooling of UPB polaritons and hot excitons and a decrease in the UPB population. On the other hand, these interactions lead to delocalization of the excitons photoexcited below the free exciton energy level. Since delocalized excitons relax to the LPB bottom much faster than do localized excitons, additional generation of hot excitons and electrons leads to a significant (by one to two orders of magnitude) increase in the population of the LPB bottom at helium temperatures.

Acceleration of the energy relaxation of polaritons by means of additional overbarrier photoexcitation leads to a strong increase in the population of states at the LPB bottom and sharply decreases the barrier for stimulated parametric scattering of polaritons excited at an LPB bending point. Therefore, additional illumination can be used as a control mechanism over polariton–polariton scattering.

ACKNOWLEDGMENTS

The authors are grateful to N.A. Gippius, L.V. Keldysh, S.T. Tikhodeev, and V.B. Timofeev for fruitful discussions, and to M.S. Skolnick for useful discussions and preparation of MC samples.

This study was supported by the Russian Foundation for Basic Research (project no. 03-02-17000), the INTAS Foundation (grant no. 01-832), and the Ministry of Science and Education of the Russian Federation (project no NSh-2164-2003.2).

REFERENCES

1. C. Weisbuch, M. Nishioka, A. Ishikawa, and Y. Arakawa, Phys. Rev. Lett. **69**, 3314 (1992).
2. F. Tassone, C. Piermarocchi, V. Savona, *et al.*, Phys. Rev. B **56**, 7554 (1997).
3. A. I. Tartakovskii, M. Imam-Ismail, R. M. Stevenson, *et al.*, Phys. Rev. B **62**, R2283 (2000).
4. M. Muller, J. Bleuse, and R. Andre, Phys. Rev. B **62**, 16886 (2000).
5. G. Malpuech, A. Kavokin, A. Di Carlo, and J. J. Baumberg, Phys. Rev. B **65**, 153310 (2002).
6. D. N. Krizhanovskii, A. I. Tartakovskii, A. V. Chernenko, *et al.*, Solid State Commun. **118**, 583 (2001).
7. A. I. Tartakovskii, D. N. Krizhanovskii, G. Malpuech, *et al.*, Phys. Rev. B **67**, 165302 (2003).
8. D. S. Citrin, Phys. Rev. B **47**, 3832 (1993).
9. V. Srinivas, J. Hryniewicz, Yung Jui Chen, and C. E. C. Wood, Phys. Rev. B **46**, 10193 (1992).
10. F. Tassone, C. Piermarocchi, V. Savona, *et al.*, Phys. Rev. B **53**, R7642 (1996).
11. F. Tassone and Y. Yamamoto, Phys. Rev. B **59**, 10830 (1999).
12. R. M. Stevenson, V. N. Astratov, M. S. Skolnick, *et al.*, Phys. Rev. Lett. **85**, 3680 (2000).
13. A. I. Tartakovskii, D. N. Krizhanovskii, and V. D. Kulakovskii, Phys. Rev. B **62**, R13298 (2000).
14. J. J. Baumberg, P. G. Savvidis, R. M. Stevenson, *et al.*, Phys. Rev. B **62**, R16247 (2000).
15. A. I. Tartakovskii, D. N. Krizhanovskii, D. A. Kurysh, *et al.*, Phys. Rev. B **65**, 081308 (2002).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 100, No. 1, 2005, pp. 139–151.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 1, 2005, pp. 156–170.

ELECTRONIC PROPERTIES OF SOLIDS

Transformation of Dielectric Properties and Appearance of Relaxation Behavior in $Pb_5(Ge_{1-x}Si_x)_3O_{11}$ Crystals

A. A. Bush[a,*], K. E. Kamentsev[a], and R. F. Mamin[b]

[a] *Moscow State Institute of Radioengineering, Electronics, and Automation (Technical University), pr. Vernadskogo 78, Moscow, 119454 Russia*

[b] *Zavoĭskiĭ Physicotechnical Institute, Kazan Scientific Center, Russian Academy of Sciences, Sibirskiĭ trakt 10/7, Kazan, 420029 Tatarstan, Russia*

**e-mail: abush@ranet.ru*

Received July 5, 2004

Abstract—The special features of the dielectric properties and conduction of ferroelectric crystals of $Pb_5(Ge_{1-x}Si_x)_3O_{11}$ ($0 \le x \le 0.67$) solid solutions were studied. Permittivity anomalies close to the temperatures $T_1 \approx 260$ K and $T_2 \approx 130$ K, the appearance of relaxator behavior at $x > 0.35$, and critical behavior of the concentration dependences of dielectric and pyroelectric characteristics at $x_1 = 0.35$ and $x_2 = 0.60$ were observed and studied. These phenomena were found to be related to the dynamics of charge localization on defects with activation energies of $U_{a1} \approx 0.6$ eV and $U_{a2} \approx 0.23$ eV. Relaxator behavior appears when the Curie point lies in the temperature region of thermal charge localization. The concentration dependence features at x_1 and x_2 are explained by the coincidence of the Curie point and the centers of the temperature regions of charge localization on the U_{a1} and U_{a2} defect levels, respectively. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Lead germanate crystals $Pb_5Ge_3O_{11}$ are uniaxial ferroelectrics with the phase transition from the $P3/m$ nonpolar phase to the $P3$ polar phase at $T_{C0} = 450$ K [1–3]. These crystals are promising as pyroelectric sensors, an optical material for recording and reading holograms, and a material for creating ferroelectric memory elements [4–9]. Thanks to the well-defined ferroelectric phase transition, which lies in the temperature region convenient for studies, lead germanate has been extensively used in the past three decades for studying soft mode problems, the appearance of the central peak, etc. [1–9]. These crystals have been model objects for studying ferroelectric phase transitions.

An important direction in studies of ferroelectric lead germanate is the preparation and investigation of various solid solutions on its base. Data on the influence of isostructural atomic substitutions on crystal properties contribute to elucidating the special features of the nature of dielectric properties and the appearance of ferroelectricity. They make it possible to change various crystal characteristics over wide ranges. Ferroelectric solid solutions that exhibit relaxator behavior are of special interest to science and technology. We found that the addition of silicon to lead germanate gradually turned it into a relaxator ferroelectric. The main general distinguishing feature of such ferroelectrics in the phase transition region is the presence of a strongly smeared maximum of the temperature dependence of permittivity and low-frequency permittivity dispersion [10–22]. Although we still use the term "phase transition," the permittivity maximum in relaxator ferroelectrics does not correspond to the real phase transition into the ferroelectric state, and no long-range order then appears. The smeared phase transition in relaxator ferroelectrics occurs as a gradual rearrangement of the crystal structure and the formation of an inhomogeneous state. The correlation radius changes with temperature but does not become infinite anywhere. For instance, X-ray studies have failed to detect structure distortions related to the phase transition to the low-temperature phase.

In recent years, ferroelectric-relaxators have been extensively studied. Nevertheless, a clear understanding of the physical mechanisms and an unambiguous interpretation of the observed phenomena are still lacking [10–22]. First, relaxator ferroelectrics were treated as objects with fluctuations of the local phase transition temperature caused by chemical disorder [10]. Later, the idea was introduced of the superparaelectric state of relaxators [11]. Recent years have witnessed the development of approaches some of which are related to treating relaxator ferroelectrics as objects with "glass behavior" [12, 13], while the others develop phenomenological models in which the system is divided into nanoregions under the action of random fields [14–16]. Recently [17–20], the conclusion was drawn that one of the conditions for the appearance of relaxator behavior is the coincidence of the phase transition region with the temperature region of thermal charge localization on defects. The possibility of shifting the Curie point in the $Pb_5(Ge_{1-x}Si_x)_3O_{11}$ (PGSO) system with respect to

the regions of charge localization on defects by changing the concentration of silicon makes this system a very convenient object for verifying this conclusion.

Several works were concerned with lead germanate-silicate crystals $Pb_5(Ge_{1-x}Si_x)_3O_{11}$ in which part of germanium atoms were replaced by silicon [3, 4, 23–31]. The main trends of variations in the ferroelectric and other properties of the system caused by the replacement of germanium with silicon were described. The T_m temperature corresponding to maximum permittivity decreased from 450 K at $x = 0$ to 220 K at $x = 0.75$ as the content of silicon in PGSO crystals increased. In crystals with $x > 0.25$, the phase transition was smeared and relaxator behavior was observed. In addition to those in the phase transition temperature region, permittivity anomalies in PGSO crystals were observed over the temperature intervals 210–270 K and 40–150 K; the corresponding permittivity maxima were close to the characteristic temperatures 260 and 130 K [29]. As the composition with $x_1 = 0.39$ was approached, the concentration dependences of the properties of solid solutions first exhibited a decrease in the permittivity ε_m in the Curie point region. The permittivity then sharply increased and passed a well-defined maximum at $x_1 = 0.39$, and the phase transition became smeared. Similar behavior was observed for compositions close to $x_2 = 0.67$. The concentration dependences of the dielectric loss tangent [29] and the pyroelectric constant [24] of solid solutions were also nonmonotonic. Note that, in [24, 29], the crystals were ascribed the composition of the batch mixtures from which they were grown.

It is pertinent to note that the dielectric, conduction, and pyroelectric properties of PGSO crystals have obviously been studied insufficiently. In particular, the low-temperature permittivity anomalies of PGSO crystals were only studied at one frequency of 1 kHz; the temperature dependences of their conduction were reported for limited frequency, temperature, and concentration ranges; the pyroelectric constants of $Pb_5Ge_3O_{11}$ crystals reported by various authors were substantially different [4–6, 24]; data on the pyroeffect in the region of permittivity anomalies at low temperatures were virtually absent; and the concentration dependences of the pyroeffect of solid solutions [24] were determined in fairly large concentration steps ($\Delta x = 0.15$). The incompleteness and contradictory character of the data impede the determination of the nature and reasons for the appearance of the low-temperature permittivity anomalies, special features of its concentration dependence, and the origin of relaxator behavior of the system. The nature of the specified dielectric property characteristics of the system of solid solutions under consideration has not been elucidated thus far.

The influence of various factors, such as the measuring field frequency, the degree of monodomainization of crystals and their calcining at various temperatures in various gas media, composition deviations from stoichiometry, etc., on the low-temperature permittivity anomalies of $Pb_5Ge_3O_{11}$ crystals was studied in [32]. The results led the authors to conclude that these anomalies were relaxation in character and related to the dynamics of thermal localization of charge carriers on defect centers with the formation of local polarized states. It was shown in [32] that, in conformity with the theoretical results described in [17–20], an analysis of the dielectric and other properties of $Pb_5Ge_3O_{11}$ crystals should be performed bearing in mind that they belonged to the class of ferroelectric-semiconductors [33–36] in which the interaction of the electronic and lattice subsystems noticeably influenced dielectric properties. The energy gap width of $Pb_5Ge_3O_{11}$ crystals is $E_g = 3.0$ eV [35]. In these crystals at $T > 300$ K, impurity conduction related to the presence of local defect levels with activation energies of 0.62 eV in the forbidden band prevails [36, 37].

When we began these studies, we suggested that the low-temperature permittivity anomalies of isostructural PGSO solid solution crystals had the same nature as those of pure lead germanate. The behavior of the dielectric properties of the system, however, has certain special features which manifest themselves as striking peculiarities of concentration dependences and the appearance of relaxator behavior. We therefore deemed it of interest to study the dielectric and conduction properties of the whole series of solid solutions in detail.

The purpose of this work was to determine the character and reason for the appearance of low-temperature permittivity anomalies in PGSO solid solutions of all concentrations and explain changes in the dielectric properties of solid solutions caused by changes in their composition, the presence of striking peculiarities of the concentration dependences, and the appearance of relaxator behavior in the system. To solve these problems, we performed detailed studies of the dielectric and conduction properties of the PGSO system over a wide temperature range at various frequencies. We took advantage of the possibility of shifting the Curie point T_C with respect to the region of thermal localization of charges by changing the concentration of silicon, which, according to [17–20], should substantially influence the dielectric properties. The permittivity anomalies were characterized in more detail by studying the temperature dependences of the pyroelectric effect of crystals at low temperatures.

2. THE PREPARATION OF CRYSTALS AND METHODS FOR STUDYING THEM

Single crystals of the composition $Pb_5(Ge_{1-x}Si_x)_3O_{11}$ were grown by cooling a molten $5PbO \cdot 3(1-y)GeO_2 \cdot 3ySiO_2$ ($0 \le y \le 1$) batch mixture at a rate of 5.0 K/h in platinum crucibles as described in [29]. Transparent light-brown crystals had an isometric habitus with a

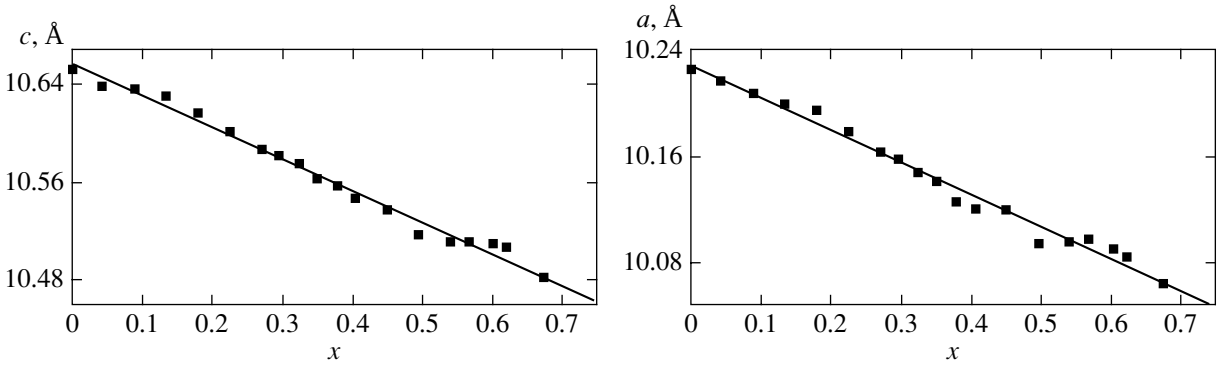


Fig. 1. Concentration dependences of the *a* and *c* parameters of the hexagonal unit cell of $Pb_5(Ge_{1-x}Si_x)_3O_{11}$ solid solution crystals.

{10$\bar{1}$0}- and {0001}-type faceting, their size across reached 4–8 mm. The chemical composition of the crystals was determined by X-ray spectral microanalysis on a CAMEBAX-301 spectrometer. The content of silicon in the crystals was found to vary within the limits $0 \le x \le 0.67$. The ratio between the concentration of silicon atoms in the crystals *x* and their concentration in the initial batch mixture *y* was 0.91 ± 0.03 over the entire range of *y* variations. Clearly, this *x*/*y* ratio equals the distribution coefficient of silicon during growing solid solution crystals from their melts. These data on the distribution coefficient can be used to refine the real concentrations of solid solution crystals studied earlier, which were assigned the compositions of the initial melts [3, 4, 23–31]. We were unable to prepare crystals with silicon concentrations $x > 0.67$, and the $x = 0.67$ concentration is likely limiting for the replacement of germanium with silicon in $Pb_5(Ge_{1-x}Si_x)_3O_{11}$ solid solutions.

The phase composition of the samples and the orientation of crystallographic axes in the crystals were determined using a DRON-4 diffractometer and a POLAM L-213M polarization microscope. The X-ray powder patterns of the crystals and the hexagonal unit cell parameters calculated from them corresponded to the data on $Pb_5(Ge, Si)_3O_{11}$ crystals obtained in [3, 4, 26, 29]. In agreement with these data, crystal symmetry did not change as the content of silicon increased. Simultaneously, the *a* and *c* hexagonal unit cell parameters decreased monotonically, without noticeable jumps (Fig. 1). These results are evidence of the absence of obvious morphotropic phase transitions in the system.

Electrophysical properties were studied for plates cut from the single crystals normally to the *c* axis. Electrodes were deposited on the base planes of the plates using a silver paste. The thickness of the plates and the area of the electrodes were $d = 0.5$–1.5 mm and $S \approx 6$ mm^2, respectively. The permittivity ε, dielectric loss tangent $\tan\delta$, and specific electrical conductivity ρ were measured using an E7-14 immittance meter or a P5083 alternating current bridge in low measuring fields. The characteristic rate of temperature variations was 5.0 K/min. The pyroelectric constant γ was determined by the quasi-static method using a V7-30 electrometer. To obtain monodomain crystals, the samples were polarized by cooling to temperatures above the Curie point and simultaneously applying a constant electric field of 0.25 kV/cm.

3. RESULTS

We studied the temperature dependences of permittivity ε, dielectric loss tangent $\tan\delta$, and specific conductivity ρ of $Pb_5(Ge_{1-x}Si_x)_3O_{11}$ crystals for various silicon concentrations over a wide temperature range of 4.2–850 K at various frequencies from 0.1 to 100 kHz. We also measured the pyroelectric constant γ over the temperature range 100–350 K. The measurement results are shown in Figs. 2 and 3. Special attention was paid to two silicon concentration regions, $0.30 < x < 0.45$ and $0.55 < x < 0.67$. It will be shown below that, at these concentrations, the effective Curie temperature $T_C(x)$ is in the temperature region of dielectric peculiarities close to the critical temperatures $T_1 \approx 260$ K and $T_2 \approx 130$ K. The results of measurements allow us to distinguish several anomalies of the temperature and concentration dependences of the dielectric, conduction, and pyroelectric properties of the crystals under consideration.

Permittivity anomalies in the regions of T_m, T_1, and T_2. In agreement with the data obtained previously, the temperature dependences of the permittivity of PGSO crystals contain a well-defined maximum ε_m at the temperature T_m in the region of the Curie point. Apart from this peak, we observe less pronounced maxima ε_1 and ε_2 close to the temperatures $T_1 \approx 260$ K and $T_2 \approx 130$ K below T_m (see Fig. 2). These maxima are accompanied by anomalies of the temperature dependences of the dielectric loss tangent $\tan\delta(T)$, which has a peak at about 230 K and a very broad asymmetric peak at 60–120 K.

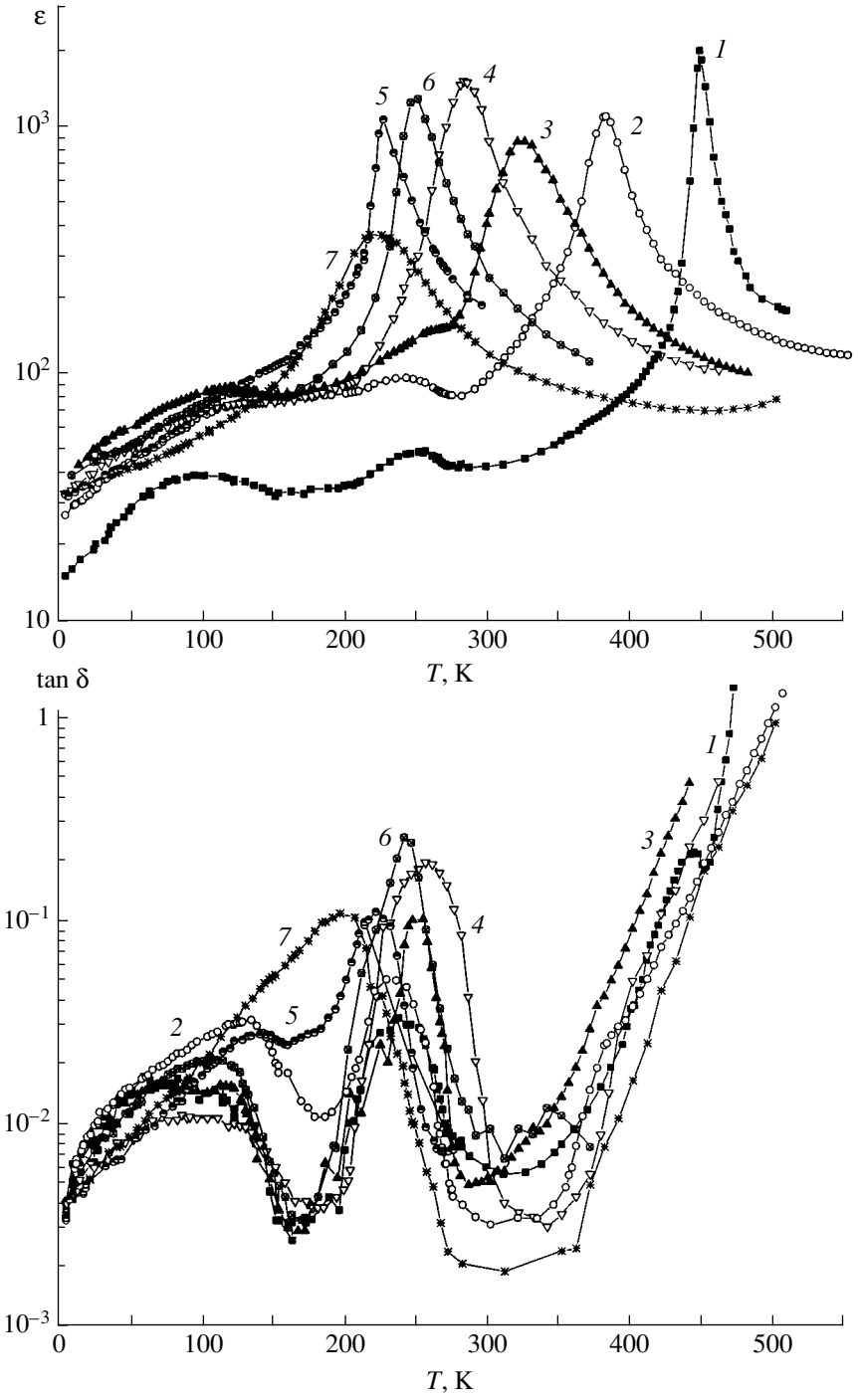


Fig. 2. Temperature dependences of permittivity ε and dielectric loss tangent $\tan\delta$ measured along the *c* axis of $Pb_5(Ge_{1-x}Si_x)_3O_{11}$ crystals at x = (*1*) 0, (*2*) 0.135, (*3*) 0.27, (*4*) 0.35, (*5*) 0.54, (*6*) 0.60, and (*7*) 0.675 at an f = 1 kHz frequency.

The concentration dependences of the T_m, T_1, and T_2 temperatures corresponding to the permittivity maxima, the maximum permittivity values themselves, the dielectric loss tangents at T_m, T_1, and T_2 (ε_m, ε_1, ε_2 and $\tan\delta_m$, $\tan\delta_1$, $\tan\delta_2$, respectively), and the concentration dependence of the maximum pyroelectric constant value γ_m close to the Curie point are shown in Fig. 4.

The position of the ε_m peak close to the Curie point changes as the content of silicon in the crystals increases and shifts toward lower temperatures from 450 K at $x = 0$ to 225 K at $x = 0.67$ (Figs. 2, 4). The position of the ε_m peak shifts to the lower temperature $T_1 \approx$ 260 K as the content of silicon increases to $x = x_1 =$ 0.35; two permittivity anomalies at T_m and T_1 then coa-

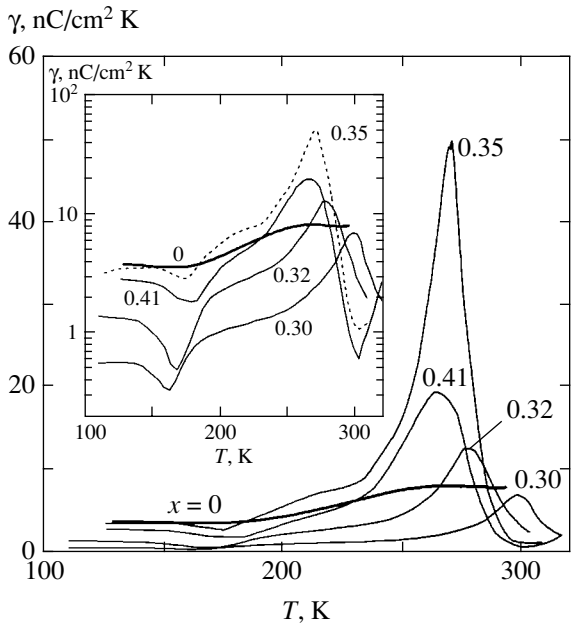


Fig. 3. Temperature dependences of the pyroelectric constant measured along the polar axis of $Pb_5(Ge_{1-x}Si_x)_3O_{11}$ crystals (numbers at curves are silicon contents). Inset shows log plots of the same dependences.

lesce ($T_m = T_1$), see Figs. 2 and 4. The permittivity peak heights ε_m and dielectric loss tangents $\tan\delta_m$ close to T_m decrease as the concentration of silicon in the crystals increases from $x = 0$ to 0.35. When the $x = 0.35$ silicon concentration is reached (when two anomalies at T_m and T_1 coalesce), the peak permittivity value ε_m increases jumpwise. The ε_m peak permittivity value for the crystals close to the Curie point decreases as x increases above x_1 and then again increases and passes the second maximum at $x_2 = 0.60$. At the same critical silicon concentration values x_1 and x_2, maxima of the concentration dependences of the peak dielectric loss tangent and pyroelectric constant values are observed (Fig. 4).

An increase in the content of silicon comparatively weakly influences the positions of the low-temperature permittivity and dielectric loss peaks. The ε_1 and $\tan\delta_1$ peaks shift to higher temperatures by approximately 20 K, and their heights increase (Figs. 2 and 4). The positions and heights of the ε_2 and $\tan\delta_2$ peaks in the region of T_2 change insignificantly.

To summarize, x_1 is the critical concentration of silicon at which two dielectric property anomalies at T_m and T_1 coalesce. This is accompanied by a jump increase in the peak permittivity value ε_m. The ε_m value decreases as the content of silicon increases further and again reaches a maximum at the next critical silicon concentration $x_2 = 0.60$. The concentration dependences of the peak dielectric loss tangent and pyroelectric constant values have maxima at the same critical silicon concentrations x_1 and x_2, which are also characterized by a noticeable broadening of the peak of the temperature dependence of permittivity in the phase transition region (Fig. 5).

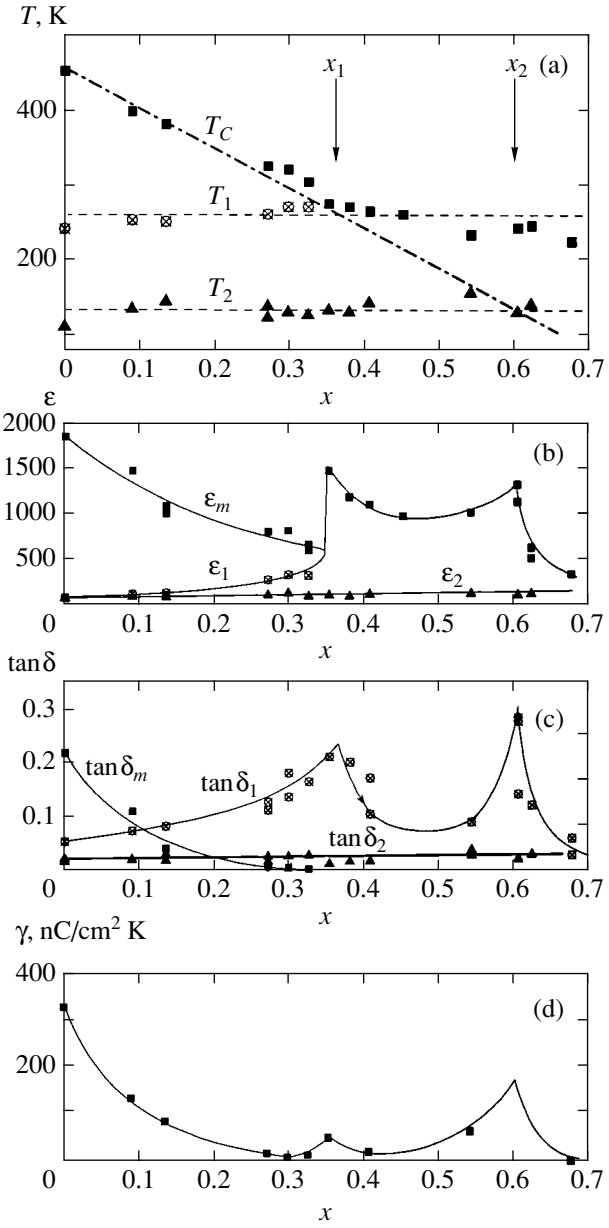


Fig. 4. Concentration dependences in the $Pb_5(G_{1-x}Si_x)_3O_{11}$ system of (a) position and (b) amplitude of permittivity maxima; (c) maximum dielectric loss tangent values in the vicinity of the temperatures T_m, T_2, and T_1 and (d) maximum pyroelectric constant values in the region of the Curie point T_m. The permittivity anomaly values were obtained by measuring the $\varepsilon(T)$ and $\tan\delta(T)$ dependences at a 1 kHz frequency; the extrapolation curve (1) (see text) is shown by the dot-and-dash line.

Dispersion of the low-temperature permittivity anomalies. The temperature dependences of permittiv-

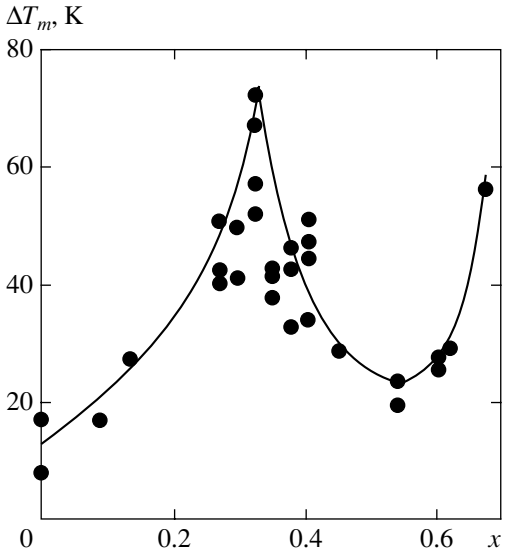


Fig. 5. Concentration dependence of the width at half-height ΔT_m of the permittivity maxima in the region of T_C in $Pb_5(Ge_{1-x}Si_x)_3O_{11}$ crystals according to $\varepsilon(T)$ dependence measurements at a 1 kHz frequency.

ity and dielectric loss tangent measured for three silicon concentrations (x = 0, 0.32, and 0.41) at various frequencies are shown in Fig. 6. The temperatures corresponding to the low-temperature $\tan\delta_1$ peak at $x < 0.35$ and the peak at $x > 0.35$ that remains after two peaks at T_m and T_1 coalesce noticeably increase as the measuring field frequency grows (Fig. 6). This is evidence of the relaxation character of these features. For instance, the temperature corresponding to the $\tan\delta_1$ peak of the crystal with x = 0 shifts from 200 to 260 K as the frequency increases from 0.1 to 100 kHz. An unusual form of permittivity dispersion is observed for the solid solutions with $x > 0.35$. Whereas the $\tan\delta(T)$ peak shifts to higher temperatures by approximately 60 K as the frequency increases from 0.1 to 100 kHz, the frequency-induced shift of the $\varepsilon(T)$ anomaly is virtually unnoticeable (Fig. 6c). At x = 0.41, the $\tan\delta_m$ peak related to the ferroelectric transition is unobservable, whereas the temperature corresponding to the $\tan\delta_1$ relaxation peak increases from 210 K at 0.1 kHz to 272 K at 100 kHz. Dispersion is observed at low frequencies not characteristic of the lattice subsystem. This is evidence that PGSO crystals are relaxator ferroelectrics.

The low-temperature ε_2 and $\tan\delta_2$ peaks at T_2 also shift to higher temperatures as the measuring field frequency increases. At x = 0, the $\tan\delta_2$ peak shifts from 116.7 K at 0.1 kHz to 144.1 K at 10 kHz (see Fig. 6a).

Transition smearing. An increase in the content of silicon in the crystals causes ferroelectric phase transition smearing, which sharply increases as the silicon concentration approaches x_1 = 0.35. The solid solutions with $x > 0.35$ acquire features characteristic of relaxator ferroelectrics. They are characterized by an obviously smeared phase transition, a sharp increase in the half-width of the permittivity maximum (see Fig. 5), and noticeable permittivity dispersion in the region of the Curie point at low frequencies.

At the silicon concentration x = 0.32, which is lower than the critical value x_1, an additional low-temperature peak ε_1 at T_1 is clearly seen against the background of the broadened main peak ε_m (Fig. 6b). This peak has a strong frequency dependence. At the concentration x = 0.41, which is higher than the critical concentration x_1, we observe the temperature and frequency dependences of permittivity with one smeared maximum and pronounced dispersion at low frequencies (Fig. 6c), as is typical of relaxators.

Nonergodicity. The heights and shapes of the permittivity and dielectric loss maxima of the crystals strongly depend on their thermal prehistory; that is, we observe nonergodic behavior. A tendency toward increasing the amplitudes of the maxima after calcining the crystals at high temperatures and decreasing these amplitudes after holding the crystals at low temperatures is noticeable. This tendency is most manifest at $x \approx x_1 = 0.35$. The form of the anomalies of the $\varepsilon(T)$ and $\tan\delta(T)$ dependences measured for crystals with silicon contents close to x_1 and x_2 changes substantially during thermal cycling. Changes in the form of these dependences caused by thermal cycling of crystals with x = 0.33 are shown in Fig. 7. Certain differences between the temperature dependences measured at different frequencies during one heating cycle and during different cycles at one frequency can be explained by the thermal cycling effect. In particular, measurements during different cycles exhibit irregularities in the sequence of decreasing the ε value depending on the frequency; no such irregularities are observed during one heating cycle.

It follows that the crystals of the system under consideration exhibit nonergodicity of the behavior of the temperature dependences of dielectric characteristics. This effect increases in the regions of compositions close to the x_1 and x_2 critical values.

The temperature dependence of resistance. To elucidate the nature of the observed features of the temperature and concentration dependences of the dielectric properties of the PGSO crystals, we studied the temperature dependences of their conduction properties. The temperature dependences of specific electrical resistance measured at various frequencies are shown in Fig. 8.

Four temperature regions which differ in the character of conduction are quite clearly seen in the temperature dependences of the specific resistance of the PGSO crystals (Fig. 8). In the high-temperature region

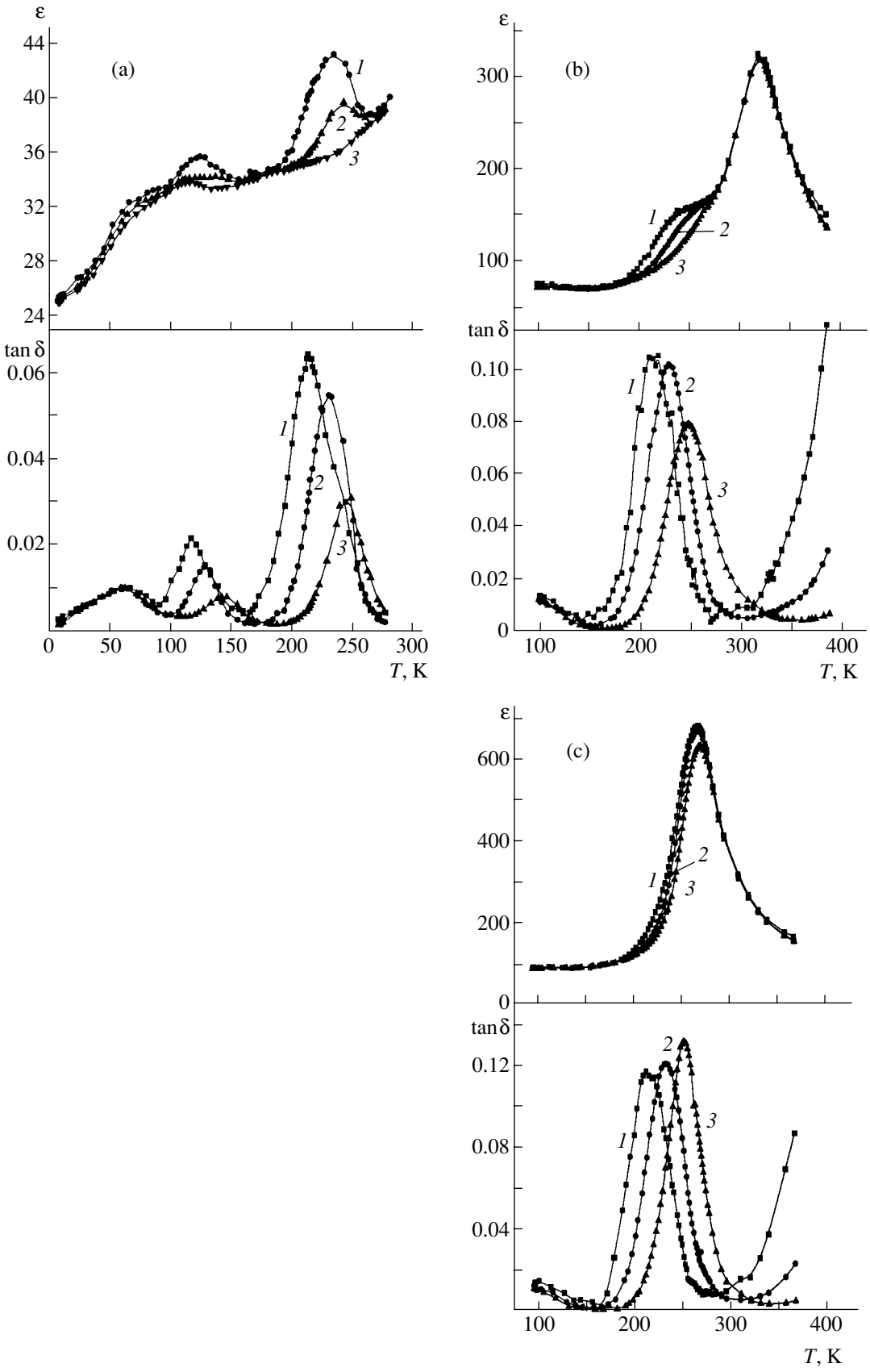


Fig. 6. Temperature dependences of permittivity ε and dielectric loss tangent tan δ measured along the *c* axis of $Pb_5(Ge_{1-x}Si_x)_3O_{11}$ crystals at *x* = (a) 0, (b) 0.32, (c) 0.41 and frequencies *f* = (*1*) 0.1, (*2*) 1, and (*3*) 10 kHz.

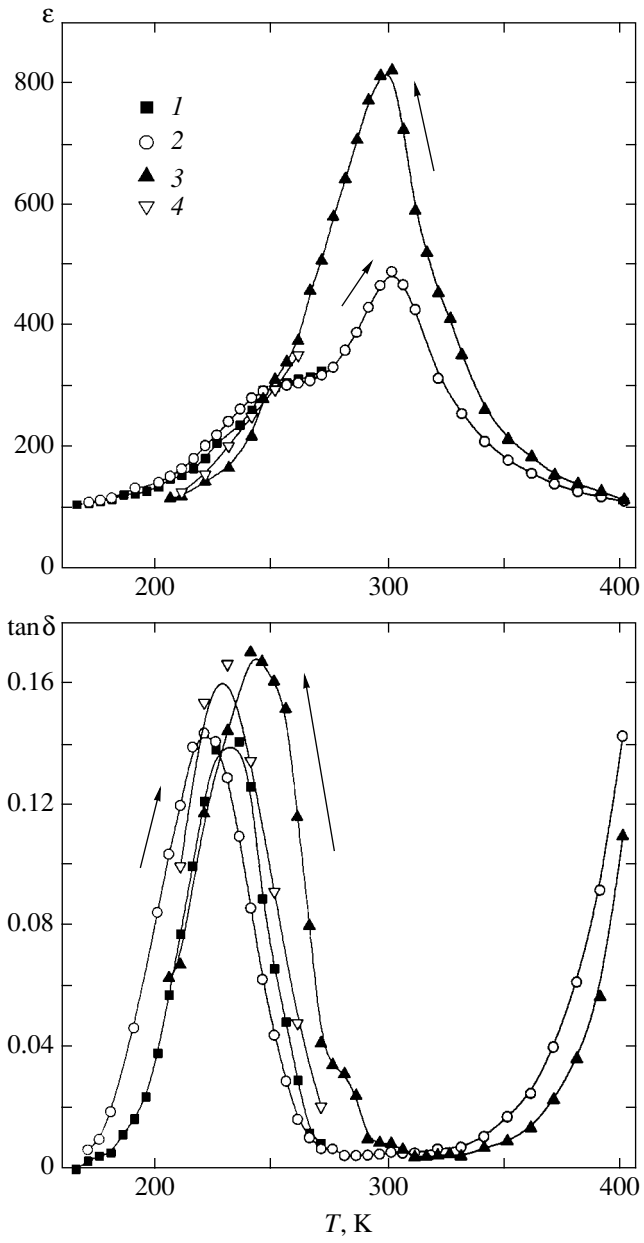


Fig. 7. Temperature dependences of permittivity ε and dielectric loss tangent $\tan\delta$ measured at a 1 kHz frequency along the *c* axis of $Pb_5(Ge_{1-x}Si_x)_3O_{11}$ crystals with x = 0.32. The results of four cycles of measurements performed one after another under (*1*) heating, (*2*) cooling, (*3*) heating, and (*4*) cooling conditions are shown. Arrows indicate direction of temperature variation.

(region I, $T > 285$ K), resistance is independent of frequency and decreases as the temperature increases according to the Arrhenius law,

$$\rho = \rho_0 \exp(U_{a1}/k_B T),$$

where k_B is the Boltzmann constant and U_{a1} is the activation energy, $U_{a1} \approx 0.64$ eV. In temperature region II (T = 210–285 K), the temperature dependence of resistance has a transient shape, resistance depends on temperature uncharacteristically, and a well-defined frequency dependence is observed. In low-temperature region III (T = 150–210 K), resistance again weakly depends on the frequency and decreases as the temperature increases following the Arrhenius law, $\rho = \rho_0 \exp(U_{a2}/k_B T)$, with the activation energy $U_{a2} \approx 0.23$ eV. In low-temperature region IV ($T < 140$ K), resistance weakly depends on the temperature and again strongly depends on the frequency.

The obvious activation dependence of conduction in temperature regions I and III (Fig. 8) is evidence of usual semiconductor conduction with thermal activation of charge carriers from the local levels U_{a1} and U_{a2} to the conduction band. The nonactivation form of the temperature dependences and the strong frequency dependence in temperature regions II and IV are evidence of a transient conduction mechanism with the predominance of hopping conduction over states localized at the U_{a1} and U_{a2} levels [38]. A decrease in resistance in these regions as the temperature decreases is indicative of the localization of charges on defects; that is, the number of charge carriers that for quite a long time reside on the corresponding U_{a1} and U_{a2} levels and can participate in hopping conduction increases as the temperature decreases. As a result, resistance decreases, which is evidence that regions II and IV are the regions of the effective thermal localization of charges. The temperatures 285 and 150 K are therefore the upper limits of the temperatures at which the effective thermal localization of charges on the levels U_{a1} and U_{a2}, respectively, begins. Note that the observed activation energies U_{a1} and U_{a2} are much smaller than the energy gap width (3.0 eV) and correspond to the energy intervals between the bottom of the conduction band and the defect levels.

Changes in the content of silicon in the crystals weakly influence their conduction properties (in particular, the positions of the temperature regions of charge localization, where thermally activated conduction transforms into hopping conduction) and the activation energies of conduction U_{a1} and U_{a2}.

The temperature dependence of pyroeffect. Peaks in the region of the Curie point characteristic of ferroelectric phase transitions are observed in the temperature dependences of the pyroelectric constant of polarized solid solution crystals (see Fig. 3). The γ_m peak height decreases as the content of silicon in the crystals increases from $x = 0$ to 0.33. At x = 0.33–0.35, γ_m increases and reaches a maximum at $x = 0.35$. It then decreases and again increases to the second local maximum at $x = 0.60$ (see Fig. 3). The temperature dependences of the pyroelectric constant below the Curie

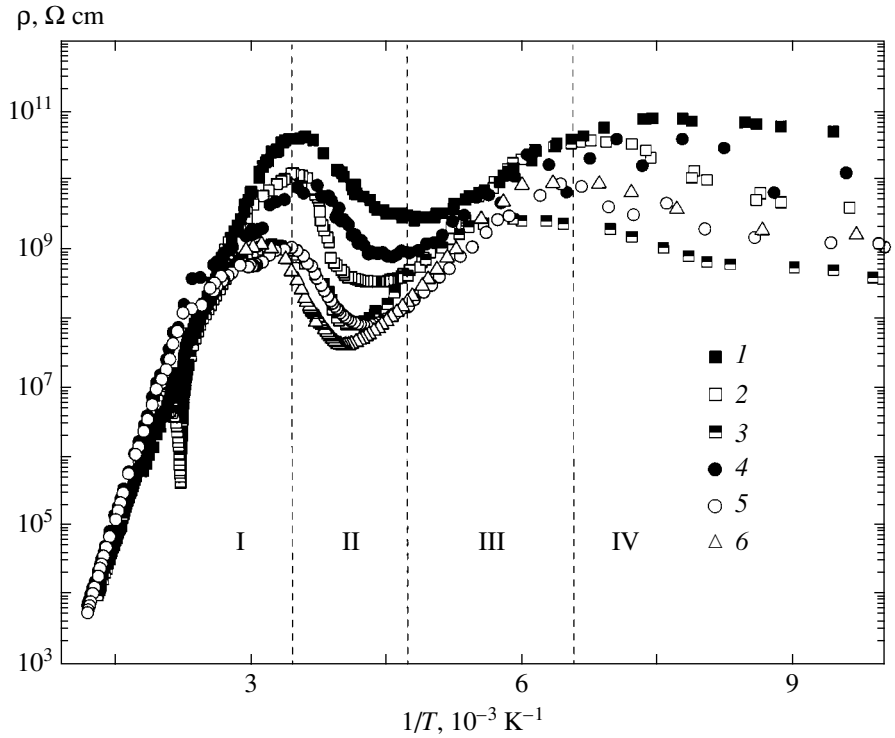


Fig. 8. Temperature dependences of specific resistance ρ measured along the *c* axis of $Pb_5(Ge_{1-x}Si_x)_3O_{11}$ crystals with x = (*1*, *2*) 0, (*3*–*5*) 0.32, and (*6*) 0.41 at frequencies f = (*3*) 0.1, (*1*, *4*, *6*) 1, and (*2*, *5*) 10 kHz.

point contain a minimum in the temperature region 170–185 K (see Fig. 3).

4. ANALYSIS AND DISCUSSION

4.1. The Effective Curie Temperature of Solid Solution Crystals

The well-defined permittivity maxima observed in the crystals of all compositions at the highest temperature $T_m(x)$ for each concentration are related to the ferroelectric phase transition or the relaxator-type smeared phase transition that occurs in the crystals. This is substantiated by the presence of a well-defined pyroelectric constant maximum in the same temperature region and the absence of the pyro effect at the higher temperatures.

The permittivity maximum at T_m shifts to lower temperatures as the content of silicon in the crystals increases. At x = 0–0.20, when a sharp phase transition occurs, the $T_m(x)$ temperature of the permittivity maximum corresponds to the Curie point $T_C(x)$ of the system. In this region of concentrations, the Curie point of the system linearly depends on the concentration of silicon, as is characteristic of many ferroelectric solid solutions,

$$T_C(x) = T_{C0} - Ax, \quad (1)$$

where T_{C0} = 450 K, A = 530 K, and x is the concentration of silicon.

At higher silicon concentrations at which the crystals acquire features of relaxator ferroelectrics, the positions of the smeared high-temperature ε(T) maxima do not correspond to a phase transition. Generally, phase transitions can be observed in them only after applying an external constant electric field [21, 22]. The effective Curie temperature of the corresponding solid solution crystals can be determined by extrapolating dependence (1) over the entire region of silicon concentrations (see Fig. 4). This dependence will be used below to analyze the special features of the behavior of the dielectric properties of the system under consideration over the entire concentration range.

The lowering of the Curie point and, accordingly, permittivity and dielectric loss peak values in the vicinity of the Curie point as the content of silicon in the crystals increases is explained by the lower polarizability of silicon cations compared with germanium cations that they substitute.

4.2. The Relaxation Character of Low-Temperature Permittivity Anomalies and the Nature of their Appearance

The low-temperature permittivity anomalies observed at $T < T_m$ are relaxation in character. This follows from the noticeable displacement of the corre-

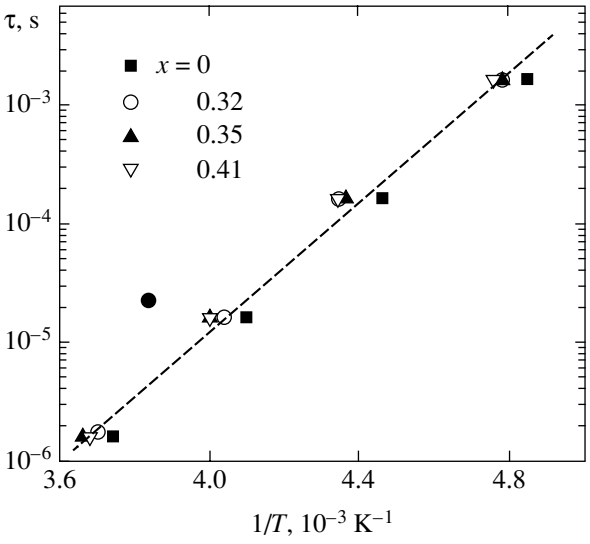


Fig. 9. Dependence of relaxation time τ on inverse temperature at various compositions.

sponding $\tan\delta(T)$ dependence peaks to higher temperatures as the measuring field frequency increases. We determined the characteristic relaxation times τ at various temperatures on the assumption that the equality $\omega\tau = 1$ ($\omega = 2\pi f$) was valid at the $\varepsilon''(T, \omega)$ maximum [39, 40]. The relaxation times determined this way are the most probable. The $\tau(T)$ dependence that we obtained was well approximated by the Arrhenius law (Fig. 9),

$$\tau = \tau_0 \exp(-U_a^\varepsilon / k_B T).$$

The activation energy U_{a1}^ε and the characteristic relaxation time τ_{01} for the permittivity anomalies in the region of T_1 were determined graphically. These values changed insignificantly when silicon atoms were introduced into the crystals and equaled $U_{a1}^\varepsilon = 0.60$ eV and $\tau_{01} = 2.0 \times 10^{-14}$ s. Analogous calculations performed for the dielectric anomalies in the vicinity of T_2 yield $U_{a2}^\varepsilon = 0.24$ eV and $\tau_{02} = 4.7 \times 10^{-14}$ s.

The peak widths of the $\tan\delta(T)$ dependences in the vicinity of 230 K were estimated by the Debye equation [39, 40]

$$\varepsilon^* - \varepsilon_\infty = \frac{\varepsilon_s - \varepsilon_\infty}{1 + i\omega\tau}$$

and the Arrhenius equation with the use of the τ_0 and U_a^ε values found above. (Here, $\varepsilon^* = \varepsilon' - i\varepsilon''$, with ε_∞ and ε_s denoting the high-frequency and static contributions to permittivity.) The resulting values were two to three times smaller than the experimentally observed peak widths of the corresponding dependences. It follows that the relaxation processes observed in the crystals were characterized by diverse relaxation times.

Analysis of the experimental temperature dependences of the conduction of the PGSO crystals performed above shows that the characteristic temperatures T_1 and T_2 of the low-temperature permittivity anomalies are in the temperature region of the thermal localization of charges on the defect levels U_{a1} and U_{a2}. The closeness of the activation energy of relaxation processes determined from dielectric measurements ($U_{a1}^\varepsilon = 0.6$ eV and $U_{a2}^\varepsilon = 0.24$ eV) to the activation energy of conduction at 300–450 K ($U_{a1} \approx 0.64$ eV) and 150–210 K ($U_{a2} \approx 0.23$ eV), as well as the temperature sequence of changes in the activation energy of conduction and the appearance of dielectric property peculiarities, is evidence that dielectric relaxation processes and the temperature dependence of conduction are determined by the same defect levels. This and the results obtained in [17–20, 32] lead us to conclude that the low-temperature permittivity anomalies in the PGSO solid solution crystals of all compositions are related to the thermal localization of charge carriers on defect centers. Filling defect levels as the temperature decreases causes the appearance of local deformations and local electric fields around charges localized on defects, which in turn stimulates the appearance of induced local polarization [19, 41]. The dynamics of these processes manifests itself in the form of low-frequency relaxation dielectric anomalies in the temperature region of the thermal filling of local centers [17–20]. This conclusion also corresponds with the low-frequency range of the dispersion of the observed dielectric anomalies. The characteristic time of induced polarization changes and, accordingly, the frequency range of permittivity dispersion are determined by the vibrational properties of local states formed around local centers and depend on the characteristic times of the lattice and electronic subsystems [17–20]. The low-frequency range of dispersion is related to the characteristic times of the thermal activation of charges from defect levels, which are much longer than the characteristic lattice times.

The weak dependence of the characteristic temperatures of the low-temperature permittivity anomalies in PGSO on concentration is explained by the weak influence of the substitution of silicon for germanium in the crystals on their main semiconductor properties that determine the appearance of these anomalies, that is, on the positions of the temperature regions of the localization of charge carriers on the U_{a1} and U_{a2} levels and level depths.

4.3. The Appearance of Relaxator Behavior in $Pb_5(Ge_{1-x}Si_x)_3O_{11}$

As the content of silicon increases in the solid solutions under consideration, the sharp ferroelectric phase

transition is replaced by a smeared transition of the relaxator type. The maximum smearing of the ferroelectric phase transition is observed for the compositions whose effective Curie temperature $T_C(x)$ lies in the temperature region of charge localization. This behavior conforms to the recently suggested phenomenological model of relaxator ferroelectrics [18–20]. According to this model, relaxator behavior arises from the relation between the lattice subsystem and the dynamics of the thermal localization of charges in traps. According to one of the conditions for the appearance of relaxator behavior, the phase transition temperature should lie in the temperature region of the thermal localization of charge carriers on defects.

The charges remain localized on defects fairly long as the temperature decreases, which fixes the local polarization direction on them [41]. The inhomogeneous polarization state formed around a charged defect is retained equally long. This inhomogeneous state effectively prevents the formation of homogeneous polarization in the system. As the Curie point is reached, spontaneous polarization has different directions in nanoregions about various defects. The system then acquires an inhomogeneously polarized state, and the phase transition becomes smeared. On the other hand, the characteristic polarization relaxation times in such a system are related not only to the dynamics of the lattice subsystem, but also to the dynamics of charge delocalization from defects. For this reason, we observe dispersion at frequencies that are much lower than lattice frequencies.

The shape of the $\varepsilon(T)$ and $\tan\delta(T)$ dependence anomalies of the crystals with the compositions in the transient region with $x \approx x_1$ substantially changes under thermal cycling (see Fig. 7). According to [17], such nonergodic behavior can be explained by dependence on the thermal prehistory of the crystal.

To summarize, we see that relaxator behavior appears when the effective Curie temperature $T_C(x)$ [see Eq. (1)] lies in the temperature region of the thermal localization of charges. This conclusion is in agreement with what follows from the theoretical model considered in [17–20].

4.4. The Nature of Defect Centers

The influence of silicon content variations on the conduction properties of the crystals is weak, which means that the effect of silicon on the electronic subsystem is insignificant. It does not create additional charge localization levels and does not change the energy depth of such levels. This leads us to conclude that the localization centers are likely related to the sublattice of lead atoms. The broad distribution of relaxation times related to localization on the U_{a2} level (see above) is evidence that this level is in reality a system of defect levels with a fairly broad energy distribution.

Some conclusions on the nature of defect centers in PGSO can be drawn from the results of the photorefractometric studies of $Pb_5Ge_3O_{11}$ [8, 9, 42, 43]. These results show that the characteristic defects in $Pb_5Ge_3O_{11}$ that play the role of charge traps in the forbidden band are part of Pb^{2+} cations transformed to Pb^{3+} under the action of optical radiation. Clearly, further detailed studies of the characteristics of defect centers in PGSO crystals will be of great importance for a more in-depth interpretation of the special features of their dielectric properties.

4.5. An Analysis of Nonmonotonic Concentration Dependences

As follows from the X-ray data (see above), the nonmonotonic character of changes in the properties of the system of solid solutions under consideration in the regions of the $x_1 = 0.35$ and $x_2 = 0.60$ concentrations cannot be explained by morphotropic phase transitions caused by concentration changes. An increase in the content of silicon in the crystals does not change their symmetry, and the hexagonal unit cell parameters then decrease monotonically without noticeable jumps (see Fig. 1).

The nonmonotonic character of the concentration dependences of the dielectric and pyroelectric solid solution characteristics in the neighborhoods of $x_1 = 0.35$ and $x_2 = 0.60$ can be explained by the influence of the localization of charge carriers on defect centers on dielectric and other properties and an increase in this influence when the Curie point gets into the region of thermal localization. According to [17–20], the shape of dielectric property anomalies should depend on the position of the Curie point relative to the characteristic temperature of the region of the thermal localization of charge carriers.

The positions of the temperature regions of charge localization close to T_1 and T_2 change insignificantly as the content of silicon in the crystals increases, whereas the effective Curie temperature $T_C(x)$ substantially decreases. For this reason, the $T_C(x)$ temperature, which initially, at $x < 0.35$, lies above the region of thermal localization, decreases as the content of silicon increases and, at $x_1 = 0.35$ and $x_2 = 0.60$, occurs in the regions of thermal localization at the U_{a1} and U_{a2} levels, respectively. Clearly, the crystals of the specified compositions, at which $T_C(x)$ is inside the thermal localization regions, should have extreme dielectric and pyroelectric characteristic values. Changes in the dielectric properties of the solid solutions close to $x_1 = 0.35$ and $x_2 = 0.60$ are critical in character. It follows that the nonmonotonic changes observed in the concentration dependences of the dielectric and pyroelectric characteristics of the crystals at about $x_1 = 0.35$ and $x_2 = 0.60$ are related to the coincidence of the effective Curie temperature [see (1)] and the temperature region of the

thermal localization of charges on the U_{a1} and U_{a2} levels, respectively, at these concentrations.

5. CONCLUSIONS

Our study of the dielectric properties and conduction of $Pb_5(Ge_{1-x}Si_x)_3O_{11}$ solid solutions revealed peculiarities in the behavior of the measured parameters caused by changes in the temperature and concentration. These were: the presence of anomalies of the temperature dependences of dielectric properties close to the characteristic temperatures T_1 and T_2 in addition to the usual anomalies close to the Curie point; the presence of well-defined maxima of the concentration dependences of dielectric properties close to the concentrations $x_1 = 0.35$ and $x_2 = 0.60$; the strong dependence of dielectric properties on the temperature prehistory of the samples close to the concentration x_1; and the appearance of relaxator behavior in the crystals at $x > 0.35$ (x is the concentration of silicon).

A combined analysis of the dielectric properties and conduction of PGSO family crystals shows that the observed dielectric property peculiarities are related to an increase in the strength of interrelation between the lattice and electronic subsystems when the region of the structural phase transition coincides with the region of the thermal localization of charges on defects. The form of the temperature–frequency dependences of permittivity anomalies in the vicinity of the T_1 and T_2 temperatures is evidence that these anomalies are relaxation in character. The appearance of these anomalies is caused by the influence of the localization of charges on defects with activation energies $U_{a1} = 0.64$ eV and $U_{a2} = 0.23$ eV on the dielectric properties of the crystals. The conduction data, according to which the T_1 and T_2 temperatures lie in the temperature regions of charge localization on the corresponding defects, lend support to this conclusion. This conclusion is also substantiated by the coincidence of the activation energies of relaxation processes determined from dielectric measurements and the activation energies of conduction and by the correspondence between the low-frequency dielectric relaxation range and the characteristic times of changes in the concentration of charges on defect levels. These results are in agreement with the theoretical concepts [17–20] that describe the influence of charge localization on defects on dielectric properties.

The reason for the pronounced anomalies of the concentration dependences of dielectric characteristics is changes in the position of the effective Curie temperature $T_C(x)$ with respect to the T_1 and T_2 temperatures of the thermal localization of charges on the U_{a1} and U_{a2} defect levels. The positions of the temperature regions of the thermal localization of charges and activation energy values remain almost unchanged as the content of silicon in the crystals increases. This is evidence that the defect centers are related to the lead sublattice. On the other hand, an increase in the content of silicon in the crystals strongly decreases the effective Curie temperature $T_C(x)$. At the concentrations x_1 and x_2, the effective Curie temperature $T_C(x)$ lies inside the regions of the thermal localization of charges T_1 and T_2. As a result, the influence of the dynamics of charge localization on dielectric properties sharply increases at these concentrations, and we observe concentration dependence maxima and relaxator behavior in agreement with the theoretical predictions [17–20]. The strong dependence of dielectric properties on the temperature prehistory of the crystals with $x \approx 0.35$ is caused by a substantial temperature-induced change in the concentration of localized charges and the retention in memory of the inhomogeneous charge distribution over traps created when crystals are held close to $T_C \sim T_1$ for a long time.

To summarize, we studied the temperature-frequency and concentration dependences of the dielectric properties and conduction of $Pb_5(Ge_{1-x}Si_x)_3O_{11}$ family crystals. Analysis of the results obtained enabled us to reveal the nature of the appearance of low-temperature dielectric property peculiarities and relaxator behavior in the system.

ACKNOWLEDGMENTS

This work was financially supported by the Russian Foundation for Basic Research (project no. 02-02-17798).

REFERENCES

1. H. Iwasaki, K. Sugii, T. Yamada, and N. Niizeki, Appl. Phys. Lett. **18**, 444 (1971).
2. S. Nanamatsu, H. Sugiyama, K. Doi, and Y. Kondo, J. Phys. Soc. Jpn. **31**, 616 (1971).
3. H. Iwasaki, S. Miyazawa, H. Kiyomada, *et al.*, J. Appl. Phys. **43**, 4907 (1972).
4. A. A. Bush and Yu. N. Venevtsev, *Monocrystals with Ferroelectric and Related Properties in the PbO–GeO$_2$ System and Possible Ranges of Their Application* (NIITÉKhIM, Moscow, 1981) [in Russian].
5. C. R. Jones, N. Show, and A. W. Vere, Electron. Lett. **8**, 346 (1972).
6. R. Watton, C. Smith, and G. R. Jones, Ferroelectrics **14**, 719 (1976).
7. T. Li and S. T. Hsu, Integr. Ferroelectr. **34**, 1495 (2001).
8. S. Mendricks, X. Yue, R. Pankrath, *et al.*, Appl. Phys. B **68**, 887 (1999).
9. H. J. Reyher, M. Pape, and N. Hausfeld, J. Phys.: Condens. Matter **13**, 3767 (2001).
10. G. A. Smolenskiĭ, V. A. Isupov, A. I. Agranovskaya, and S. N. Popov, Fiz. Tverd. Tela (Leningrad) **2**, 2906 (1961) [Sov. Phys. Solid State **2**, 2584 (1961)].
11. L. E. Cross, Ferroelectrics **76**, 241 (1987).
12. R. Blinc, J. Dolinsek, A. Gregorovic, *et al.*, Phys. Rev. Lett. **83**, 424 (1999).

13. E. V. Colla, L. K. Chao, and M. B. Weissman, Phys. Rev. Lett. **88**, 017601 (2002).
14. A. E. Glazounov and A. K. Tagantsev, Phys. Rev. Lett. **85**, 2192 (2000).
15. V. Westphal, W. Kleemann, and M. D. Glinchuk, Phys. Rev. Lett. **68**, 847 (1992).
16. W. Kleemann, Int. J. Mod. Phys. B **7**, 2569 (1993).
17. R. F. Mamin, Pis'ma Zh. Éksp. Teor. Fiz. **58**, 534 (1993) [JETP Lett. **58**, 538 (1993)].
18. R. F. Mamin, Fiz. Tverd. Tela (St. Petersburg) **43**, 1262 (2001) [Phys. Solid State **43**, 1314 (2001)].
19. R. F. Mamin, Izv. Ross. Akad. Nauk, Ser. Fiz. **67**, 1157 (2003).
20. R. F. Mamin, S. A. Migachev, S. I. Nikitin, *et al.*, Ferroelectrics **296**, 157 (2003).
21. R. Sommer, N. K. Yushin, and J. J. van der Klink, Phys. Rev. B **48**, 13230 (1993).
22. G. A. Samara, J. Phys.: Condens. Matter **15**, R367 (2003).
23. W. Eysel, R. W. Wolfe, and R. E. Newnham, J. Am. Ceram. Soc. **56**, 185 (1973).
24. S. Yu. Stefanovich, V. V. Chechkin, V. D. Sal'nikov, and Yu. N. Venevtsev, Kristallografiya **21**, 540 (1976) [Sov. Phys. Crystallogr. **21**, 300 (1976)].
25. A. G. Belous, V. V. Dem'yanov, and Yu. N. Venevtsev, Fiz. Tverd. Tela (Leningrad) **19**, 1694 (1977) [Sov. Phys. Solid State **19**, 1578 (1977)].
26. W. Muller-Lierheim, W. Gebhardt, H. H. Otto, and G. Busse, Ferroelectrics **20**, 299 (1978).
27. E. V. Sinyakov, V. V. Gene, and A. Ya. Kreĭcherek, Fiz. Tverd. Tela (Leningrad) **21**, 1223 (1979) [Sov. Phys. Solid State **21**, 709 (1979)].
28. M. Polomska, M. Malinowski, and H. H. Otto, Phys. Status Solidi A **56**, 335 (1979).
29. A. A. Bush and Yu. N. Venevtsev, Kristallografiya **26**, 349 (1981) [Sov. Phys. Crystallogr. **26**, 198 (1981)].
30. K. Takagashi, S. Shirasaki, K. Takamatsu, *et al.*, Jpn. J. Appl. Phys., Suppl. **22** (2), 73 (1983).
31. A. A. Bush and E. A. Popova, Fiz. Tverd. Tela (St. Petersburg) **46**, 875 (2004) [Phys. Solid State **46**, 902 (2004)].
32. A. A. Bush, K. E. Kamentsev, M. V. Provotorov, and T. N. Trushkova, Fiz. Tverd. Tela (St. Petersburg) **46**, 1668 (2004) [Phys. Solid State **46**, 1722 (2004)].
33. V. M. Fridkin, *Ferroelectric Semiconductors* (Nauka, Moscow, 1976; Consultants Bureau, New York, 1980).
34. V. T. Gabriélyan, P. V. Ionov, K. A. Mikhaĭlova, and O. A. Arikalov, Kristallografiya **19**, 176 (1974) [Sov. Phys. Crystallogr. **19**, 106 (1974)].
35. V. V. Mikhnevich, A. V. Mikulenok, and A. V. Myasoedov, Fiz. Tverd. Tela (Leningrad) **27**, 932 (1985) [Sov. Phys. Solid State **27**, 572 (1985)].
36. A. Mansingh, K. N. Srivastava, and B. Singh, J. Appl. Phys. **50**, 4319 (1979).
37. L. E. Cross and T. W. Cline, Ferroelectrics **11**, 333 (1976).
38. N. F. Mott and E. A. Davis, *Electronic Processes in Non-Crystalline Materials* (Clarendon, Oxford, 1971; Mir, Moscow, 1974).
39. H. Frolich, *Theory of Dielectrics: Dielectric Constant and Dielectric Loss*, 2nd ed. (Clarendon Press, Oxford, 1986; Inostrannaya Literatura, Moscow, 1960).
40. V. Daniel, *Dielectric Relaxation* (Academic, London, 1967).
41. A. P. Levanyuk, V. V. Osipov, A. S. Sigov, and A. A. Sobyanin, Zh. Éksp. Teor. Fiz. **76**, 345 (1979) [Sov. Phys. JETP **49**, 176 (1979)].
42. X. Yue, S. Mendricks, Y. Hu, *et al.*, J. Appl. Phys. **83**, 3473 (1998).
43. X. Yue, S. Mendricks, T. Nikolajsen, *et al.*, J. Appl. Phys. **86**, 1186 (1999).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 100, No. 1, 2005, pp. 14–21.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 1, 2005, pp. 18–26.

ATOMS, MOLECULES, OPTICS

Electron-Forbidden Energy Gap of Hydrogen in a Wide Pressure Interval¶

A. G. Khrapak[a] and K. Yoshino[b]

[a]*Institute for High Energy Density, Russian Academy of Sciences, Moscow, 125412 Russia*
[b]*Department of Electronic Engineering, Graduate School of Engineering, Osaka University, 565-0871, Osaka, Japan*
e-mail: khrapak@mail.ru

Received June 24, 2004

Abstract—A simple model is used for estimating the bottom energy of the electron conduction band and the electron-forbidden gap energy. It is shown that electrons in liquid hydrogen are localized not in electron bubbles, as was considered previously, but in molecular negative ions surrounded by voids about 0.5 nm in radius. The conductivity of fluid hydrogen at not very high pressures is connected to transfer of positively charged clusters and negatively charged bubbles. As the pressure and density increase, molecular dissociation occurs and electron localization on atoms becomes more favorable, also with the creation of a void around atomic negative ions. At a sufficiently high concentration of atoms, the probability of tunnel transition of an electron from one atom to another becomes close to unity, the energy level of the negative ion degenerates in the band, and the conductivity is caused by the transfer of these quasifree electrons. It is supposed that this charge transfer mechanism may play an important role in the region of fluid hydrogen metallization. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Investigations of hydrogen behavior in the solid, liquid, and plasma state are of significant importance from both a scientific and technological standpoint. They are important in astrophysics because hydrogen is the most abundant chemical element, which constitutes about three-fourths of the matter in the Universe. Jupiter and Saturn together contain over 400 Earth masses, most of which is hydrogen, heated and compressed to high temperatures and pressures. Hydrogen is fluid under these conditions. The convective motion of electrically conducting hydrogen produces the magnetic field of giant planets by dynamo action (see [1] and references therein). Knowledge of the equation of state and physical properties of hydrogen and its isotopes is very important for successful solution of the problem of inertial nuclear fusion. An intriguing possibility of metastable metallic, and even superconducting, phases of solid hydrogen at ambient pressure has been predicted [2].

In 1935, Wigner and Huntington predicted that molecular diatomic hydrogen would undergo a transition to a metallic state at an imposed pressure of about 25 GPa [3]. Current predictions are in a range close to 300 GPa [4]. However, despite unrelenting assault of experiments, dense solid hydrogen shows no evidence of metallic behavior [5]. In a fluid, electric conductivity measurements under multiple-shock compression indicate that hydrogen becomes metallic at pressures of about 140 GPa, nine times the initial fluid density, and a temperature of about 3000 K [6–10]. Electric conductivity has also been measured under single-shock compression up to 20 GPa and 4600 K [11]. Those experiments show that conductivity is thermally activated similarly to the semiconducting fluid, and becomes greater than 1 Ω^{-1} cm^{-1} at 200 GPa and 400 K. The pressure dependence of the conductivity measured in [9] is shown in Fig. 1. The change in slope at

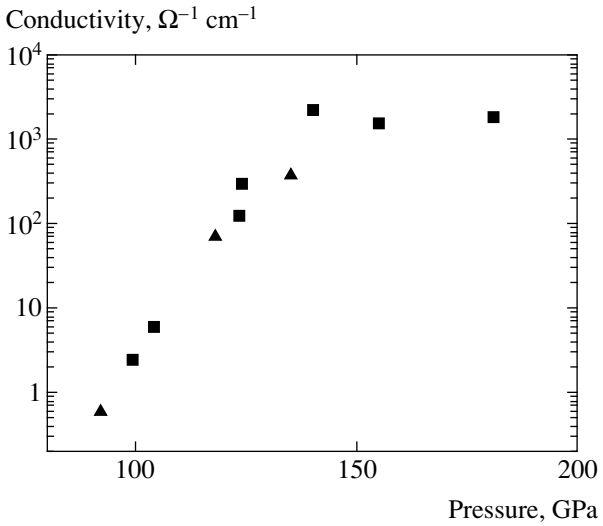


Fig. 1. Electrical conductivity of H_2 and D_2 as a function of pressure [9]. The change in slope at 140 GPa is the transition from semiconducting to metallic fluid.

¶ This article was submitted by authors in English.

140 GPa is indicative of the transition to the metallic state. An analysis of the measurements in the range 93–120 GPa (semiconducting regime) resulted in the equation typical of liquid semiconductors,

$$\sigma = \sigma_0 \exp[-E_g(\rho)/2k_BT], \quad (1)$$

where σ is the electric conductivity, σ_0 weakly depends on the density ρ, $E_g(\rho)$ is the density-dependent electron bandgap of the fluid, k_B is the Boltzmann constant, and T is the temperature. If the temperature dependence of conductivity is related to transition of electrons from the ground state to the continuous spectrum, then E_g coincides with the ionization potential of the hydrogen molecule in matter. The results of the least-square fit of the experimental data to Eq. (1) are

$$\begin{aligned} E_g(\rho) &= 1.22 - 62.6(\rho - 0.300), \\ \sigma_0 &= 90, \quad 0.290 \le \rho \le 0.319, \end{aligned} \quad (2)$$

where $E_g(\rho)$ is in eV, ρ is in mol/cm^3, and σ_0 is in Ω^{-1} cm^{-1}. We note that a value 200–300 Ω^{-1} cm^{-1} is typical of liquid semiconductors [12]. The bandgap was also estimated as $E_g = 11.7 \pm 1.7$ eV at the point $\rho = 0.13 \pm 0.005$ mol/cm^3 in the single-shock experiments [11]. The metallization density is defined to be the density at which the mobility bandgap E_g is reduced by pressure to $E_g \sim k_BT$, at which point E_g is filled in by fluid disorder and thermal smearing. We have $E_g(\rho) \approx k_BT$ at the density about 0.316 mol/cm^3 and the temperature about 2600 K (0.22 eV). Thus, fluid hydrogen becomes metallic at about 140 GPa and 2600 K via continuous transition from a semiconducting to metallic fluid.

The bandgap E_g has been measured only for solid H_2 and D_2 at low temperatures (about 5 K) and low (saturation) pressures [13, 14]. The vacuum ultraviolet absorption spectra of these two hydrogen isotopes are practically identical. The low-energy component of the spectra below 15 eV was assigned to Wannier exciton transitions. The analysis of the higher order terms of the Wannier series in [14] implies that $E_g \approx 14.7$ eV in hydrogen and $E_g \approx 14.9$ eV in deuterium. These values are close to the gas phase ionization potentials of the hydrogen molecules: $I_g = 15.43$ eV for H_2 and $I_g = 15.47$ eV for D_2. Knowledge of E_g makes it possible to estimate the energy of the bottom of the electron conduction band V_0. In fact, the molecular ionization potential in dielectric matter, as follows from the close-coupling approximation, is related to the gas phase ionization potential by

$$E_g = I_g + P_+ + V_0, \quad (3)$$

where P_+ is the polarization energy of the medium by a positive ion. For estimating the value of P_+, the Born formula can be used:

$$P_+ = -\frac{e^2}{2R_i}\left(1 - \frac{1}{\varepsilon}\right), \quad (4)$$

where R_i is the radius of the cavity where a point charge resides, surrounded by a homogeneous liquid or solid with the dielectric constant ε. Usually, R_i is chosen equal to the crystallographic ionic radius or to the hard-core radius of the neutral parent molecule. Good agreement with the results of the theoretical estimates of P_+ for solid rare gases [15, 16] may be achieved for

$$R_i = R_s, \quad R_s = \left(\frac{3}{4\pi N}\right)^{1/3}, \quad (5)$$

where R_s is the radius of the cell occupied by a molecule in the medium with a concentration of molecules N. Substitution of Eq. (5) in Eq. (4) gives $P_+ = 0.7$ eV for H_2 and $P_+ = 0.8$ eV for D_2. Thus, according to this estimation, the energy of the bottom of the electron conduction band V_0 is approximately –0.05 eV in solid H_2 and +0.20 eV in solid D_2.

The sign and value of V_0 are determined by competition between the polarization and exchange interactions of an electron with molecules of the medium,

$$V_0 = T_e + P_e, \quad (6)$$

where $P_e < 0$ is the energy of the medium polarization by an electron and $T_e > 0$ is the minimum kinetic energy that a free electron can have in a system of short-range repulsive scatterers. With decreasing N, the relative contribution of the polarization interaction increases and V_0 must therefore also be negative in the liquid and gas phases of hydrogen. In an ideal gas, the optical model [17] may be used for estimation of V_0,

$$V_0 = \frac{2\pi\hbar^2}{m} LN, \quad (7)$$

where m is the electron mass and L is the electron–molecule scattering length. This implies that L must be negative and demonstrate a Ramsauer–Tawnsend minimum. The scattering length was obtained in spectroscopic investigation of the properties of electrons localized above the surface of solid hydrogen [18, 19]. The resonant energy of the electron transition between ground and excited surface states was found in this experiment to depend on the density N of H_2 molecules in the vapor phase. The linear density shift of the transition energy was interpreted in terms of the optical model (fully analogous to the Fermi shift of energy of high excited Rydberg atoms in a gas atmosphere) by means of Eq. (7). This gave the value $L = -0.14 \pm 0.04$ nm and negative V_0, which is in qualitative agreement with the previously discussed results of spectroscopic investigations of solid hydrogen. We note that the currently accepted value is $L = +0.067$ nm. Discussion of possible reasons of this discrepancy can be

found in [19]. Most probably, discrepancy occurs because practically all measurements of the electron scattering length were performed at temperatures high enough for the rotational degrees of freedom to be excited. Only Zavyalov and Smolyaninov [18, 19] performed their experiments at cryogen temperatures.

A negative value of V_0 indicates the absence of a potential barrier for penetration of an electron from the gas phase to bulk liquid or solid hydrogen. At first sight, this contradicts a number of well-known experimental facts. First, the possibility of electron localization above the surface of condensed He, Ne, and H_2 is usually connected with the existence of a potential barrier for electrons at the surface of these three matters having small polarizability of atoms or molecules [20, 21]. Second, in the experiments on mobility of charge carriers in liquid [22–24] and solid [25–27] hydrogen, a very low mobility of negative charges (on the same order as or, in some cases, even less than the mobility of positive charges) was observed. The current interpretation of this effect supposes that, just as in liquid and solid helium, the positive charges represent clusters (Atkins' snowballs [28]) consisting of a positively charged molecular ion surrounded by a layer of neutral molecules, and the negative charges in condensed hydrogen are electrons localized in bubbles or voids of several atomic sizes [21]. This interpretation also implies the existence of a sizable potential barrier of about 1–2 eV at the surface of the electron bubble. Third, irradiated solid hydrogen displays a number of interesting spectral features. Hydrogen mixtures containing tritium, when cooled below a temperature of about 10 K, show additional lines in the fundamental absorption spectrum [29]. The new lines were interpreted as Stark-shifted molecular transitions whose appearance was caused by the presence of trapped charges of both signs in the lattice as a result of the ionizing tritium radioactivity. Proton- and γ-irradiated samples show the same features. The analysis of Stark shifts resulted in the conclusion that two species of each charge exist, one mobile and one less mobile. Each of the less mobile charge species is responsible for the induced absorption features. The mobile negative charge is thought to be a small polaron, and its immobile counterpart is then an electron trapped in the form of a bubble [30]. In addition to the Stark-shifted features, a number of spectral features also attributed to trapped electrons have been observed in irradiated solid hydrogen (see, e.g., [31] and references therein). A simple square-well model for the electron bubble gives a good fit to the observed spectra only for an unreasonably large well depth $V_0 = 3.8$ eV [32].

One of the aims of the present work is to eliminate the aforementioned contradictions between different experiments. With the help of a simple model, we show that, even in the case of a negative V_0 (but not very large in absolute value), two-dimensional electron surface states may exist owing to the additional potential barrier at the surface, whose appearance is related to different ranges of polarization and exchange forces. An important role of the polarization energy at the surface and interface was also reported for other dielectric and semiconducting systems [33]. Using the fact of the recently observed formation of the H_2^- ion in solid hydrogen [34–36], we conjecture that the low mobility of negative charges in condensed hydrogen is a result of the electron capture by a hydrogen molecule and bubble creation around it, but not the result of electron bubble creation. We assume that, in the case of irradiated liquid and solid hydrogen, the availability of the admixture of hydrogen (deuterium) atoms is decisive and electrons are localized in H^- (D^-), surrounded by voids of smaller size than in the case of H_2^- (D_2^-). Near the metallization pressure of hydrogen, considerable dissociation of molecules (about 10%) occurs [8]. Electrons are localized in atomic negative ions. With increasing pressure, overlapping of the neighboring atomic negative ion states should result in formation of an extended electron band and lead to the insulator–metal transition. In this paper, results of our determination of the density dependence of V_0 and E_g are presented and compared with the results of single-shock experiments.

2. ENERGY OF THE BOTTOM OF THE ELECTRON CONDUCTION BAND

In dense fluids and solids, the interaction between atoms and molecules plays an important role, optical model (7) is inapplicable for estimation of V_0, and more detailed consideration of Eq. (6) is necessary. The energy P_e of the polarization interaction of electrons with matter differs from the ion polarization energy P_+. Calculations performed for solid rare gases [16] are well fitted by

$$P_+ = -\frac{1}{a}\varphi_+\left(\frac{\alpha}{a^2}\right),$$
$$\varphi_+(t) = \frac{3.154t - 3.860t^2}{1 + 2.55t - 4.750t^2}, \quad (8)$$

$$P_e = -\frac{1}{a}\varphi_e\left(\frac{\alpha}{a^3}\right),$$
$$\varphi_e(t) = \frac{4.966t + 0.924t^2}{1 + 3.244t + 0.957t^2}, \quad (9)$$

where $a \approx 1.2R_s$, α is the atomic or molecular polarizability, and t satisfies the conditions $0 < t < 0.2$. Approximation (8) coincides numerically with the calculation data in [15] and Born formula (4). It is well known that, for helium isotopes and other rare gases, the value of V_0 depends significantly on the atomic density and is practically independent of isotopic composition, aggregative state, and the type of crystal structure [37]. There-

fore, there is good reason to believe that Eqs. (8) and (9) can be used for solid and fluid hydrogen and deuterium. The results of estimating the polarization energy of positive ions and electrons near the triple point of H_2 and D_2 are listed in the table.

For estimating the minimum electron kinetic energy T_e, it is necessary to specify the short-range part of the interaction potential. The interaction of an electron with an atom or molecule in the vacuum can be qualitatively described by a simple model potential shown in Fig. 2 [38, 39],

$$V(r) = \begin{cases} \infty, & r \le R_c, \\ -\dfrac{\alpha e^2}{2mr^4}, & r > R_c. \end{cases} \quad (10)$$

The only unknown parameter of the potential, the solid core radius R_c, is fitted as follows. In the case where a stable negative ion of the corresponding species exists (as is the case with H^-), the value of R_c is sought with which the solution of the Schrödinger equation with potential (10) gives the correct value of the electron affinity EA. The atomic hydrogen has EA = 0.754 eV [40], which results in R_c = 0.032 nm. A negative ion of molecular hydrogen does not exist in vacuum. In this case, it is possible to use the known relation between R_c, α, and the electron scattering length L [41],

$$L = \sqrt{\frac{\alpha}{a_0}}\cot\sqrt{\frac{\alpha}{a_0 R_c^2}}, \quad (11)$$

where a_0 is the Bohr radius. Substitution of the value L = –0.14 nm in Eq. (11) results in R_c = 0.052 nm. This value of R_c seems quite reasonable because, in the scattering of two closed atoms, it should be twice the atomic R_c, but correlation in positions of the atoms in the molecule slightly reduces the value of R_c.

Knowing the radius of the molecular hard core, it is possible to calculate the value of zero-point electron energy T_e. An approach commonly used for this is based on the Wigner–Seitz model [15, 37, 42]. In this model, the medium is divided into equivalent spheres of radius R_s. Each sphere contains a hard core of radius R_c at its center. A free electron may be in any cell with equal probabilities. Therefore, the electron wavefunction $\psi(r) \sim r^{-1}\sin[k(r - R_c)]$ and $\psi'(r)$ must be continuous at the cell boundaries, which is possible only if $\psi'(R_s) = 0$. This gives

$$T_e = \frac{\hbar^2 k^2}{2m}, \quad kR_s = \tan[k(R_s - R_c)]. \quad (12)$$

The results of estimations of T_e and the values of V_0 and E_g following from Eqs. (6) and (3) are also listed in the table. It follows that our estimation of the ionization potential E_g in solid hydrogen is in good agreement with spectroscopic measurements [14] and the bottom of the electron conduction band V_0 is negative in all cases considered.

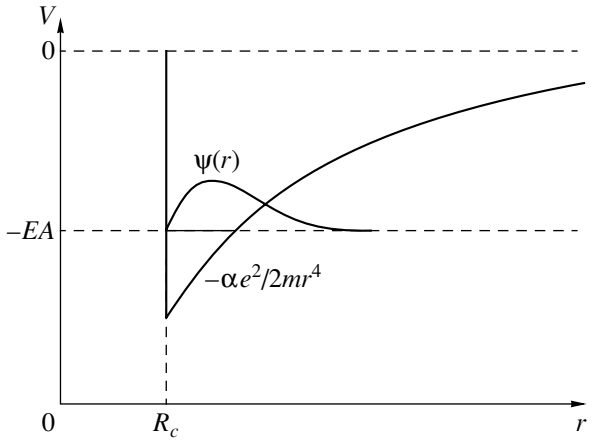


Fig. 2. A model potential for the electron–atom or electron–molecule interaction $V(r)$ and the electron wavefunction $\psi(r)$ in the negative ion.

This model makes it possible to determine the dependence of V_0 and E_g on the fluid density ρ. The results of our calculations are shown in Figs. 3 and 4. V_0 has a minimum at a density of about 0.3 g/cm^3 and becomes positive at about 0.5 g/cm^3. Such behavior is typical of all fluid rare gases having negative V_0 under ambient conditions. The forbidden energy gap decreases with density and, within the experimental errors, coincides with E_g produced by single-shock compression at the point 0.53 g/cm^3.

3. SURFACE POTENTIAL BARRIER AND LOCALIZATION OF ELECTRONS ON THE SURFACE OF HYDROGEN

For liquids with a positive value of V_0 (for example, helium and neon), the electron transfer from the vac-

Results of estimation of the characteristic energies of H_2 and D_2 near their triple points: the positive ion P_+ and the electron P_e polarization energy; the minimum electron kinetic energy T_e; the energy of the bottom of the electron conduction band V_0; and the electron mobility gap or ionization potential E_g. All values are in eV

	Hydrogen		Deuterium	
	liquid	solid	liquid	solid
P_+	–0.67	–0.78	–0.78	–0.91
P_e	–1.08	–1.26	–1.26	–1.48
T_e	0.95	1.09	1.09	1.27
V_0	–0.08	–0.16	–0.16	–0.18
E_g	14.68	14.49	14.53	14.38

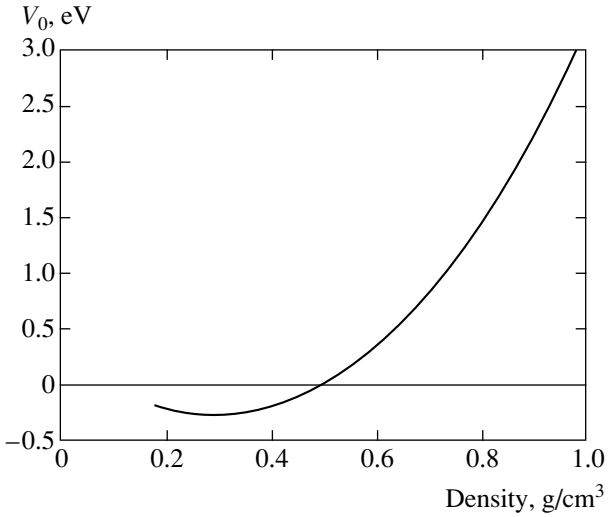


Fig. 3. Energy of the bottom of the electron conduction band V_0 in fluid D_2 as a function of density ρ.

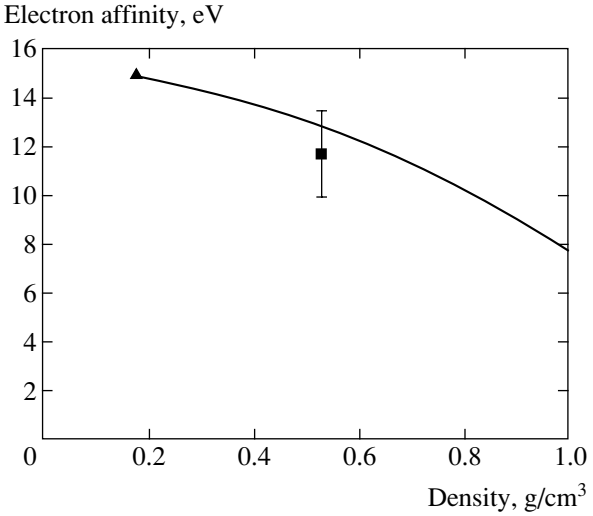


Fig. 4. Electron-forbidden energy gap in fluid D_2 as a function of density ρ.

uum into the liquid is hampered by this barrier. An electron approaching the surface from the vacuum nevertheless feels the influence of its positive image charge inside the liquid. The potential of this attractive image force above the surface is given by

$$V(z) = -\frac{Qe^2}{z}, \quad Q = \frac{\varepsilon - 1}{4(\varepsilon + 1)}, \tag{13}$$

where z denotes the coordinate perpendicular to the surface and ε is the dielectric constant of matter. The attraction by the image force and the barrier given by V_0 lead to a bound surface state [20, 21]. The electron is, however, still partially free to move along the surface and has high mobility in these directions. For liquid helium, the potential barrier $V_0 \sim 1$ eV is high in comparison with the binding energy of the localized electron. Therefore, it is possible to set $V_0 = \infty$ with good accuracy and to take the presence of the interface into account by the boundary condition of the wavefunction $\Psi|_{z=0} = 0$. The attractive potential then gives rise to a hydrogen-like wave function with the Bohr radius becoming a_0/Q. The energy levels correspond to the Rydberg series, and the electron energy spectrum is given by

$$E_n(\mathbf{k}) = \frac{\hbar^2\mathbf{k}^2}{2m} - \frac{mQ^2e^4}{2\hbar^2n^2}, \quad n = 1, 2, \ldots, \tag{14}$$

where $\mathbf{k}$ is a two-dimensional wavevector of the electron parallel to the helium surface. Owing to small polarizability of helium ($Q \approx 1/144$), the ground-state binding energy is also small ($E_1(0) \approx 7.5$ K $\ll V_0$) and the assumption $V_0 = \infty$ is quite reasonable in this particular case. The electron is localized at a distance on the order of 100 nm from the surface, and, therefore, the true behavior of the interaction potential at the distance of the interatomic order from the surface is not very important. The frequencies of the 1–2 and 1–3 transitions correspond to 125.9 and 148.6 GHz, respectively.

In the case of condensed hydrogen, the situation is different. As we have shown above, V_0 is negative in both liquid and solid hydrogen, and at first sight, the surface electron localization is impossible. Such states were nevertheless observed [18, 19, 43, 44]. We now consider the spectroscopic measurements with surface electrons on solid hydrogen surfaces in more detail [19]. A tunable laser source makes it possible to observe the photoresonance of the surface electrons when changing the potential of the lower electrode U (and, consequently, the confining electric field E) altered the electron spectrum. The photoresonance signal amplitude depended linearly on the laser intensity and on the surface charge density. The transition frequency in the limit of zero electric field E and hydrogen gas pressure P was equal to 3.15 ± 0.05 THz. As in the case of similar experiments for electrons over ^{3}He and ^{4}He [45, 46], the energy spectrum can be approximated by introducing the Rydberg correction δ into Eq. (14),

$$E_n(0) = -\frac{mQ^2e^4}{2\hbar^2(n+\delta)^2}, \tag{15}$$

where δ is independent of n. Measurements of δ for ^{3}He and ^{4}He gave –0.014 and –0.022, respectively, while δ = –0.11 was obtained for solid hydrogen [18, 19]. As we already mentioned, Zavyalov and Smolyaninov found that, for electrons over solid hydrogen and deuterium, the transition frequencies depend strongly on vapor density. Analysis of this dependence made it possible to determine the scattering length L that we use in

our estimations. That the scattering length is negative is important evidence that V_0 is negative.

To understand why creation of the localized surface states is possible in the case of negative V_0, we consider the interaction of an electron with the surface of hydrogen in more detail. The interaction potential $V(z)$ is shown schematically in Fig. 5. Inside condensed hydrogen, the potential energy of the long-range polarization interaction P_e is determined, for example, by Eq. (9). Approaching the surface, P_e increases. At the surface, it is tied continuously to the image force potential, which is determined by Eq. (13) far from the surface. An electron located just at the surface interacts with half the molecules that it interacts with in the bulk matter. It is therefore reasonable to assume that, at the surface, the polarization energy is approximately $P_e/2$. In addition to the polarization interaction, there is a short-range exchange interaction of the electron with electrons of hydrogen molecules, which results in the shift of the free electron energy V_0 by a positive value T_e. The dependence of $T_e(z)$ is significantly more abrupt than that of $P_e(z)$. We approximate it by the step function. The resulting surface potential depicted in Fig. 5 by solid line represents the potential barrier for electrons penetrating from the vacuum side. It is obvious that, if V_0 is not too small, the surface electron localized states may exist.

For determination of the surface electron energy spectrum, we use an even simpler potential. Outside hydrogen, it coincides with the image force potential (13) down to $z = R_c = 0.052$ nm. At shorter distances, the potential is considered to be constant, $V(z) = V(R_c)$. Inside hydrogen, the potential is also considered to be constant, $V(z) = V_0 - P_e/2$. Solution of the Schrödinger equation gives a spectrum in Eq. (15) with $\delta \approx -0.2$ that is practically independent of n. For a model thus simplified, the agreement with the experimental value $\delta = -0.11$ is quite satisfactory. It is worthwhile to note that we used the continuity conditions

$$\psi(-0) = \psi(+0), \quad m_{\text{eff}}^{-1}\psi'(-0) = m^{-1}\psi'(+0) \quad (16)$$

for the electron wavefunction $\psi(z)$ [47], where m_{eff} is the electron effective mass. In liquid helium, $m_{\text{eff}} \approx m$ and both masses cancel out in Eq. (16). In solid hydrogen, $m_{\text{eff}} \approx 0.2m$ [48]. The sudden change in the electron effective mass at the hydrogen surface results in a significant increase in surface electron binding energy.

4. STRUCTURE AND ENERGY SPECTRUM OF ATOMIC AND MOLECULAR NEGATIVE IONS IN CONDENSED HYDROGEN

The electron affinity to atoms and molecules increases in condensed dielectrics in comparison with its value in vacuum [49–53]. This effect was observed by Lukin and Yakovlev [49] and Sowada and Holroyd [50]

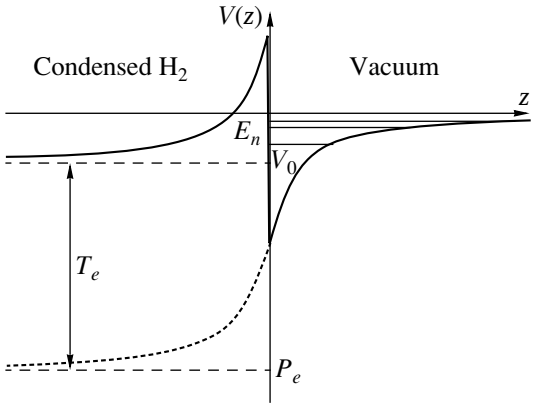


Fig. 5. Schematic arrangement of the electron interaction potential at the surface of condensed hydrogen.

in experiments on conductivity of solutions of molecular oxygen in different dielectric liquids. Such solutions were exposed to short X-ray radiation pulses. This resulted in ionization of the solvent and a sharp growth in conductivity. Then, the electrons were localized at neutral oxygen molecules with the creation of negative O_2^- ions. After this, the conductivity dropped abruptly because the mobility of heavy O_2^- ions is several orders less than the mobility of free electrons. Then, a laser pulse in the visible spectrum was produced. If the laser frequency exceeded a certain threshold value, then photodetachment occurred, accompanied by new growth of conductivity. The photodetachment energy was found to be significantly higher than the electron photodetachment energy from the oxygen molecule in vacuum (the difference was more than 1 eV). This effect is the result of a strong polarization interaction of the bound electron with atoms or molecules of the solvent. A more detailed discussion of this effect can be found in [51–53].

As a result of irradiation or thermal dissociation, a certain amount of atomic hydrogen may be present in condensed molecular hydrogen. We therefore perform our estimations for both atomic and molecular negative ions. At the moment of the electron transition to the level of a negative ion, surrounding matter can be considered undisturbed. The electron binding energy in the negative ion can then be estimated from the solution of the Schrödinger equation with a potential slightly different from that in Eq. (10) and Fig. 2; it is presented in Fig. 6. At the surface of the surrounding void, the negative ion with $R = R_s$, the interaction potential varies stepwise by a value of T_e. The estimation of the electron affinity gives, for example, EA ≈ 1.33 eV for H^- in liquid hydrogen at the triple point and EA ≈ 1.42 eV for D^- in solid deuterium, also at the triple point. A continuous red emission spectrum was observed during proton-beam irradiation of solid D_2 and H_2, maximizing near 830 nm (1.49 eV) [31, 54]. We believe that the electron attachment to D and H is responsible for this

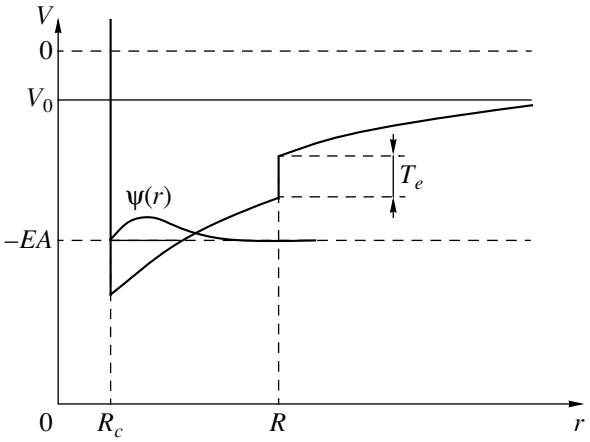


Fig. 6. Model potential for the electron–atom or electron–molecule interaction $V(r)$ and the electron wavefunction $\psi(r)$ in the negative ion inside a bubble of radius R.

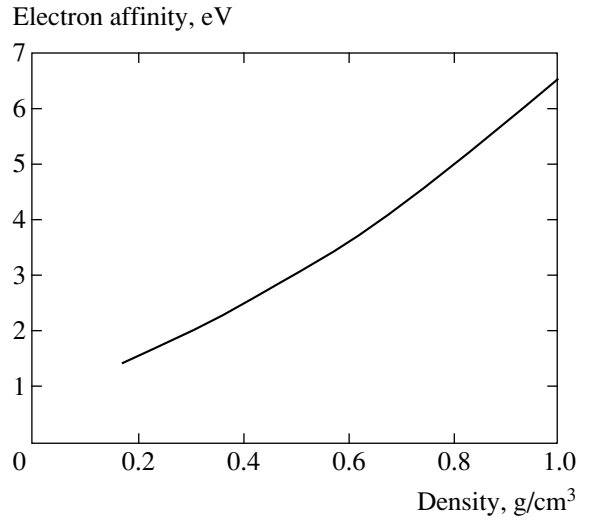


Fig. 7. Electron affinity to atomic deuterium in fluid deuterium as a function of density.

emission. Similar estimation shows that the electron affinity to the hydrogen molecule in undisturbed hydrogen is negative. Hence, the radiative formation of H_2^- and D_2^- ions is impossible. With increasing fluid density, the electron affinity increases. The results of our calculation of the electron affinity to atomic deuterium in fluid molecular deuterium are shown in Fig. 7.

After atomic negative ion formation, the interaction of its outer electron with surrounding matter results in the creation of a void around the ion, with decreasing electron energy. At the bubble radius about 0.5 nm, the electron energy shift is about 0.15–0.20 eV. With the potential barrier at the surface of the void, which is of order of $P_e/2$, the electron detachment energy is approximately equal to 1.9 eV for H^- in liquid H_2 and 2.2 eV for D^- in solid D_2. The last value is somewhat different from the experimentally measured value about 3.1 eV of the electron bound-free transition energy in proton-irradiated solid deuterium [32]. Nevertheless, we believe that photodetachment of electrons from the D^- ions may be responsible for this ultraviolet absorption spectrum and suppose that a more refined calculation of the hydrogen negative ion spectrum is capable of improving agreement with experiment. Creation of negative H_2^- and D_2^- ions also becomes possible inside voids of sufficiently large radius. For example, the electron detachment energy for the D_2^- ion in a void 0.5 nm in diameter in solid D_2 is approximately 1.3 eV. Close values are valid for H_2^- in liquid and solid H_2 and for D_2^- in liquid D_2. This allows us to suppose that low mobility of negative charges in liquid and solid hydrogen [22–27, 39] is related to electron localization in molecular or atomic (when dissociation of molecules occurs) negative ions surrounded by bubbles or voids.

5. CONCLUSIONS

A simple model for estimating the bottom energy of the electron conduction band V_0 and the forbidden energy gap E_g was proposed based on the experimental investigation of the exciton absorption spectrum in condensed hydrogen. Estimation of E_g is in good agreement with values obtained in measuring conductivity by single-shock wave experiments. It was shown that electrons in liquid hydrogen are localized not in the electron bubbles, as was considered previously, but in molecular negative ions surrounded by voids of about 0.5 nm in radius. The conductivity of fluid hydrogen at not very high pressures is related to the transfer of heavy complexes—positively charged clusters and negatively charged bubbles. With increasing pressure and density, molecular dissociation occurs and the electron localization on atoms becomes more favorable, also with the creation of a void around atomic negative ions. At a sufficiently high concentration of atoms, the probability of tunnel transition of an electron from one atom to another becomes close to unity, the energy level of the negative ion degenerates in the band, and the conductivity is caused by the transfer of these quasifree electrons. This mechanism of charge transfer may play an important role in the region of fluid hydrogen metallization.

ACKNOWLEDGMENTS

A.G.K. gratefully acknowledges support from Osaka University, the Russian Foundation for Basic Research, and Deutsche Forschungsgemeinschaft.

This work was supported by the Presidium of the Russian Academy of Sciences under the Program on Thermophysics and Mechanics of Extreme Energy Stimulation.

REFERENCES

1. W. J. Nellis, Planet. Space Sci. **48**, 671 (2000).
2. N. W. Ashcroft, Phys. Rev. Lett. **21**, 1748 (1968).
3. E. Wigner and H. B. Huntington, J. Chem. Phys. **3**, 764 (1935).
4. P. P. Edwards and F. Hensel, Nature **388**, 621 (1997).
5. R. J. Hemley and H. Mao, J. Low Temp. Phys. **122**, 331 (2001).
6. S. T. Weir, A. C. Mitchell, and W. J. Nellis, Phys. Rev. Lett. **76**, 1860 (1996).
7. V. Y. Ternovoi, A. S. Filimonov, V. E. Fortov, *et al.*, Physica B (Amsterdam) **265**, 6 (1999).
8. W. J. Nellis, S. T. Weir, and A. C. Mitchell, Phys. Rev. B **59**, 3434 (1999).
9. V. E. Fortov, V. Y. Ternovoi, S. V. Kvitov, *et al.*, JETP Lett. **69**, 926 (1999).
10. V. E. Fortov, V. Y. Ternovoi, M. V. Zhernokletov, *et al.*, JETP **97**, 259 (2003).
11. W. J. Nellis, A. C. Mitchell, P. C. McCandless, *et al.*, Phys. Rev. Lett. **68**, 2937 (1992).
12. N. F. Mott, Philos. Mag. **24**, 2 (1971).
13. A. Gedanken, B. Raz, and J. Jortner, J. Chem. Phys. **59**, 2752 (1973).
14. K. Inoue, H. Kanzaki, and S. Suga, Solid State Commun. **30**, 627 (1979).
15. L. E. Lyons and M. G. Sceats, Chem. Phys. Lett. **6**, 217 (1970).
16. A. M. Ratner, Phys. Rep. **269**, 197 (1996).
17. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory*, 3rd ed. (Nauka, Moscow, 1974; Pergamon Press, Oxford, 1977).
18. V. V. Zavyalov and I. I. Smolyaninov, JETP Lett. **44**, 182 (1986).
19. I. I. Smolyaninov, Int. J. Mod. Phys. **15**, 2075 (2001).
20. M. W. Cole, Rev. Mod. Phys. **46**, 451 (1974).
21. V. B. Shikin and Yu. P. Monarkha, *Two-Dimensional Charged Systems in Helium* (Nauka, Moscow, 1989) [in Russian].
22. B. Halpern and R. Gomer, J. Chem. Phys. **51**, 1031 (1969).
23. Y. Sakai, E. H. Böttcher, and W. F. Schmidt, J. Jpn. Inst. Electr. Eng. A **61**, 499 (1983).
24. A. A. Levchenko and L. P. Mezhov-Deglin, J. Low Temp. Phys. **89**, 457 (1992).
25. A. A. Levchenko and L. P. Mezhov-Deglin, JETP Lett. **54**, 234 (1991).
26. A. A. Levchenko and L. P. Mezhov-Deglin, Sov. J. Low Temp. Phys. **17**, 229 (1991).
27. A. B. Trusov, L. P. Mezhov-Deglin, and A. A. Levchenko, JETP Lett. **63**, 376 (1996).
28. K. R. Atkins, Phys. Rev. **116**, 1339 (1959).
29. P. C. Soers, E. M. Fearon, P. E. Roberts, *et al.*, Phys. Lett. A **77**, 277 (1980).
30. R. L. Brooks, M. A. Selen, J. L. Hunt, *et al.*, Phys. Rev. Lett. **51**, 1077 (1983).
31. J. A. Forrest and R. L. Brooks, Phys. Rev. B **55**, 906 (1997).
32. J. A. Forrest, J. L. Hunt, and R. L. Brooks, Can. J. Phys. **68**, 1247 (1990).
33. A. A. Zakhidov and K. Yoshino, Synth. Met. **64**, 155 (1994).
34. T. Miyazaki, K. Yamamoto, and Y. Aratono, Chem. Phys. Lett. **232**, 229 (1995).
35. M. C. R. Symons, Chem. Phys. Lett. **247**, 607 (1995).
36. T. Miyazaki, J. Kumagai, and T. Kumada, Radiat. Phys. Chem. **60**, 381 (2001).
37. W. F. Schmidt, *Liquid State Electronics of Insulating Liquids* (CRC Press, Boca Raton, 1997).
38. A. G. Khrapak, High Temp. **13**, 775 (1975).
39. A. V. Berezhnov, A. G. Khrapak, E. Illenberger, and W. F. Schmidt, High Temp. **41**, 425 (2003).
40. H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Academic, New York, 1957; Fizmatgiz, Moscow, 1960).
41. I. T. Iakubov and A. G. Khrapak, Phys. Rev. A **51**, 5043 (1995).
42. I. T. Iakubov and A. G. Khrapak, Rep. Prog. Phys. **45**, 697 (1982).
43. A. M. Troyanovskii, A. P. Volodin, and M. S. Khaikin, JETP Lett. **29**, 382 (1979).
44. A. M. Troyanovskii and M. S. Khaikin, Sov. Phys. JETP **54**, 214 (1981).
45. C. Grimes, T. R. Brown, M. L. Burns, and C. L. Zipfel, Phys. Rev. B **13**, 140 (1976).
46. A. P. Volodin and V. S. Edelman, JETP Lett. **30**, 633 (1979).
47. M. W. Cole, Phys. Rev. B **2**, 4239 (1970).
48. A. Gedanken, B. Raz, and J. Jortner, Chem. Phys. Lett. **14**, 326 (1972).
49. V. Lukin and B. S. Yakovlev, Chem. Phys. Lett. **42**, 307 (1976).
50. U. Sowada and R. A. Holroyd, J. Chem. Phys. **70**, 3586 (1979).
51. A. G. Khrapak and K. F. Volykhin, JETP **88**, 320 (1999).
52. A. G. Khrapak, P. Tegeder, E. Illenberger, and W. F. Schmidt, Chem. Phys. Lett. **310**, 557 (1999).
53. P. D. Grigor'ev and A. M. Dyugaev, JETP **88**, 325 (1999).
54. J. A. Forrest, R. L. Brooks, J. L. Hunt, *et al.*, Phys. Rev. B **46**, 13820 (1992).

Journal of Experimental and Theoretical Physics, Vol. 100, No. 1, 2005, pp. 152–164.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 1, 2005, pp. 171–185.

STATISTICAL, NONLINEAR, AND SOFT MATTER PHYSICS

Formation of Structures in a Dusty Ionosphere

B. A. Klumov[a], G. E. Morfill[a], and S. I. Popel[b]

[a]*Centre for Interdisciplinary Plasma Science, Max-Planck-Institut für Extraterrestrische Physik, D-85740 Garching, Germany*

[b]*Institute of Geosphere Dynamics, Russian Academy of Sciences, Moscow, 119334 Russia*

e-mail: klumov@mpe.mpg.de; s_i_popel@mtu-net.ru

Received May 28, 2004

Abstract—An analysis of the evolution of microscopic particles (dust grains) in the Earth's ionosphere and their effect on ionization in the middle atmosphere is presented. It is shown that summer conditions in the polar ionosphere, which are characterized by an ambient air temperature below 150 K and presence of supersaturated water vapor, facilitate the formation of dust structures in the middle atmosphere, such as noctilucent clouds and polar mesospheric summer echoes. The ionospheric plasma composition can change significantly in the regions occupied by these structures. Depending on photoelectric properties of the grains, their presence may lead to excess, or decrease in, electron concentration and complex behavior of ion concentration. The proposed self-consistent model of the ionosphere allows for grain growth, sedimentation, and charging and can be used to explain the behavior of ionization under summer conditions in the polar ionosphere. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Processes taking place in plasmas containing microscopic particles have been the subject of extensive studies in recent years. Systems of this kind are generally called complex (or dusty) plasmas. The particles are rapidly charged by recombination of electrons and ions on their surface, photoelectric emission, and other processes. The resulting plasmas are characterized by additional length and time scales and, therefore, by much more complex behavior, such as instabilities and waves of new types. The interest in dusty plasmas is due to their widespread occurrence in nature. Dust grains are present in the interstellar medium, planetary magnetospheres and ionospheres, and the atmospheres of comets. Frequently, dust grains strongly affect, and may even determine, environmental conditions. Understanding of the processes taking place in complex plasmas has significantly improved owing to laboratory experiments conducted since the mid-1990s [1–6]. This motivates the application of methods developed in studies of dusty plasmas to ionospheric plasmas. As shown below, the ionosphere may contain both nanometer- and micron-sized dust particles. In this paper, we call them *microparticles*, since this term is common in studies of dusty plasmas. Note also that the charging of particles greater than 10 nm in size is described by models analogous to those used in analyses of charging of micron-sized particles. However, the present study is of interdisciplinary nature, because its results can be applied in other areas, such as atmospheric physics, geophysics of micron-scale and nanostructured objects, and ecology.

Microparticles build up in the ionosphere as a result of both bombardment by micrometeorites (which subsequently burn in the upper part of the middle atmosphere, at altitudes of 80 to 100 km) and convective transport of particles from the lower atmosphere (soot produced by forest fires or particles of volcanic origin). It is commonly believed (see [7]) that these processes lead to the buildup of dust grains several nanometers in diameter at altitudes of 80 to 100 km. Their concentration varies between 10 and 1000 cm^{-3}, depending on the season and micrometeorite activity. It can be shown that the presence of these particles does not modify the state of the ionospheric plasma under normal conditions. The situation is totally different in the polar mesosphere under summer conditions at altitudes of 80 to 90 km: the ambient temperature falls below 150 K, the water vapor that is present at these altitudes supersaturates (e.g., see [8]), and dust grains grow. As their diameter reaches a certain value (at a certain time), their presence begins to determine local ionization characteristics of the ionosphere. When the micrometeoritic activity is relatively high, narrow sporadic layers of elevated ionization are frequently observed in the middle and upper atmosphere (at altitudes of 90 to 110 km) [9], where the most common ions are those of sodium, potassium, iron, and their oxides (apparently of meteoritic origin). We do not discuss the physics associated with these layers.

One important feature of the polar ionosphere under summer conditions is the presence of dust layers (very thin on the atmospheric scale) located at altitudes of 80 to 85 km (*noctilucent clouds*, or NLC) or 85 to 95 km (*polar mesospheric summer echoes*, or PMSE). NLC consist of submicron-sized particles. Their vertical optical thicknesses are much less than unity, but they can be seen by the naked eye at sunset, whereas PMSE

(apparently consisting of charged nanometer-sized particles) cannot be observed by optical methods and manifest themselves by strong radio reflections observed with radars at frequencies between 50 and 1000 MHz [10]. Strong correlation between observations of NLC and PMSE suggests that they have a common origin. In the literature, NLC and PMSE are frequently grouped together under the common term *polar mesospheric clouds* (PMC) [8]. These objects are discussed in this paper with a view to constructing a self-consistent model that can explain the most important observations.

The paper is organized as follows. In Section 2, we discuss the summer physical conditions in the polar ionosphere and present both computed spectra of solar radiation in the mesosphere and observed concentrations of electrons and ions. In Sections 3 and 4, we develop a theoretical model describing the evolution of dust structures in the middle atmosphere and discuss the results of numerical simulations of polar mesospheric clouds. Finally, we state the main conclusions.

2. SUMMER CONDITIONS IN THE POLAR IONOSPHERE

The formation of NLC and PMSE (hereinafter called dust structures[1]) takes place in the polar atmosphere at mesospheric altitudes (80–100 km) between the end of May and the end of August. In this period, the polar mesosphere is the coldest place on Earth. The ambient air temperature there falls below 150 K [8], and water vapor supersaturates. This leads to conditions favoring the growth of dust grains. The dominant nucleation mechanism appears to be the condensation of water molecules on nanometer-scale particles, which are always present at mesospheric altitudes. The characteristic grain size is a few nanometers [11], and their concentration typically is 10 to 1000 cm^{-3}. They are either products of volcanic eruptions and forest fires brought into the mesosphere from the lower atmosphere by convective mixing of the atmospheric air or (more likely) products of combustion of micrometeorites in the middle atmosphere. The micrometeorite flux toward the Earth is 100 Mt per day (e.g., see [12]). It mainly consists of centimeter-sized objects. They burn at altitudes of 80 to 100 km, supplying nanometer-sized particles to the mesosphere. It should be noted here that observation data concerning nanometer-scale particles in the upper part of the middle atmosphere are practically unavailable, because they cannot be detected by optical methods. Since information about particle composition and size distribution can be obtained only by means of sounding rocket flights [13], the quantities indicated above should be interpreted only as conjectures about the dust contained in the mesosphere [7].

[1] The term *layered* is also used in literature, because their vertical size, 10 to 1000 m, is small on the atmospheric altitude scale.

Figure 1 illustrates the summer conditions in the polar ionosphere that are important for the physics of PMCs. In particular, the figure shows vertical profiles of neutral-gas temperature, saturated water vapor pressure, water vapor pressure, and estimated microparticle concentration. Particle nucleation conditions are satisfied at altitudes between 78 and 92 km, where water vapor is supersaturated. Note that it is in this altitude range where NLC and PMSE are observed.

Under conditions of local water vapor supersaturation, the nucleation of nanometer-scale particles is followed by their charging. At the altitudes in question, the ionosphere is characterized by a complicated composition of its charged component, which consists of electrons and positive and negative ions. Positive ions can be tentatively divided into two groups. The so-called simple, or primary, ions N_2^+, O_2^+, and NO^+ are produced both by solar radiation (primarily via photoionization of nitrogen monoxide resulting in NO^+ formation) and by electron impact ionization in collisions with high-energy electrons that escape from the radiation belts. The other group contains the ions $H^+(H_2O)_n$ (so-called clusters or proton hydrates), where n is the hydration number (normally, $n \le 10$). They are produced from simple ions via a complicated chain of charge transfer reactions [14]. It is important that the two kinds of ions are characterized by substantially different recombination rate constants α_{rec}: $\alpha_{rec} \sim 10^{-7}$ cm^3 s^{-1} for simple ions, whereas $\alpha_{rec} \sim 10^{-5}$ cm^3 s^{-1} for proton hydrates. It should also be noted that the concentration of positive ions of each group is a complicated function of ionization rate q_e, ionization source (i.e., the relative contributions of solar radiation and fast electrons to ionization), as well as of the composition of the neutral middle atmosphere, particularly on the NO concentration (since photoionization primarily involves NO molecules). The composition of the positive-ion subsystem strongly depends on the neutral-gas temperature T_n. This is explained by the fact that the cluster bond strength is very small when n is large. For example, clusters with $n \ge 10$ cannot form at $T_n \ge 160$ K, because they are destroyed in collisions with neutrals at a rate faster than their formation rate. In this case, the simple positive ions and proton hydrates with small n are the dominant components.

The mesosphere may also contain negative ions [15]. However, they can be neglected at the altitudes of interest here (80–100 km). Indeed, the primary negative ion O_2^- is produced in the three-body attachment reaction

$$e + 2O_2 \longrightarrow (O_2^- + O_2)$$

(with rate constant $k_{att} \approx 10^{-31}$ cm^6 s^{-1}) and consumed in the photodetachment reaction

$$\hbar\omega + O_2^- \longrightarrow O_2 + e$$

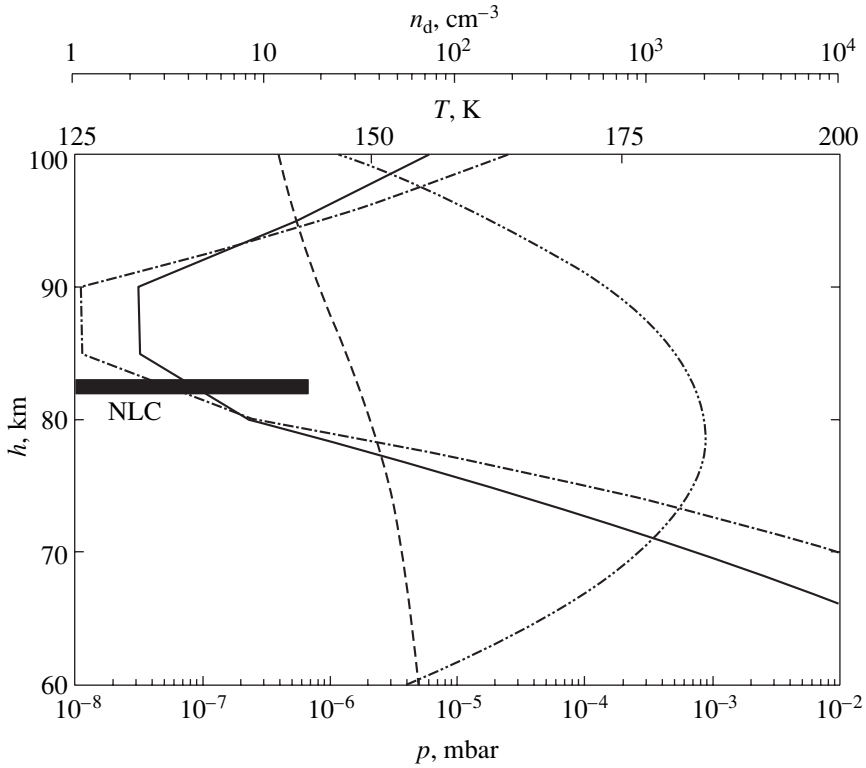


Fig. 1. Qualitative estimates for vertical profiles of air temperature (solid curve), partial pressure of water vapor (dashed curve), pressure of saturated water vapor (dash–dot curve), and nanometer-sized particle concentration (dash–dot–dot curve). Water vapor is supersaturated at altitudes of 78 to 92 km. Since the degree of saturation is $\Phi_{H_2O} = P_{H_2O}/P_{sat} \geq 1$, ice particles can form as a result of heterogeneous nucleation on nanometer-sized grains. The altitude range in which noctilucent clouds are observed is indicated.

(with rate constant $k_{photo} \geq 0.3$ s^{-1} at the altitudes in question), in the reaction with atomic oxygen

$$O_2^- + O \longrightarrow O_2 + O + e$$

(with rate constant $k_{det} \approx 3 \times 10^{-10}$ cm^3 s^{-1}), and in charge transfer reactions

$$O_2^- + M \longrightarrow M^- + O_2.$$

The fastest process is the charge transfer via collisions with water molecules. The corresponding rate constant is $k_{ct}^{H_2O} \approx 10^{-11}$ cm^3 s^{-1}.

The O_2^- kinetics in the lower ionosphere can be described by the equation [16]

$$\frac{\partial[O_2^-]}{\partial t} \approx k_{att} n_e [O_2]^2 - k_{det}[O_2^-][O] - k_{photo}[O_2^-] - k_{ct}^{H_2O}[O_2^-][H_2O], \quad (1)$$

where brackets denote species concentrations. The balance condition for the sources and sinks of O_2^- yields an estimate for the O_2^- concentration:

$$[O_2^-] \approx k_{att} n_e [O_2]^2 / (k_{photo} + k_{det}[O] + k_{ct}^{H_2O}[H_2O]) \approx 0.1\text{–}1 \text{ cm}^{-3}.$$

This quantity is substantially smaller than the characteristic concentrations of positive ions and electrons at the altitudes in question, which are estimated as follows:

$$n_e \approx n_i \approx (q_e/\alpha_{rec})^{1/2} \approx 10^2\text{–}10^4 \text{ cm}^{-3},$$

where the characteristic ionization rate in the mesosphere is $q_e \sim 0.1$–10 cm^{-3} s^{-1}.

Microscopic currents of electrons and ions in the ambient plasma and their recombination on the surface of a microparticle result in its charging. This leads to a considerable decrease in positive-ion and electron concentrations in the regions occupied by NLC and PMSE. Effects due to solar radiation may lead to much more complicated behavior of charged species in dusty ionospheric plasmas, depending on the photoelectric properties of microparticles, their concentration, and grain size. This behavior is discussed in detail below.

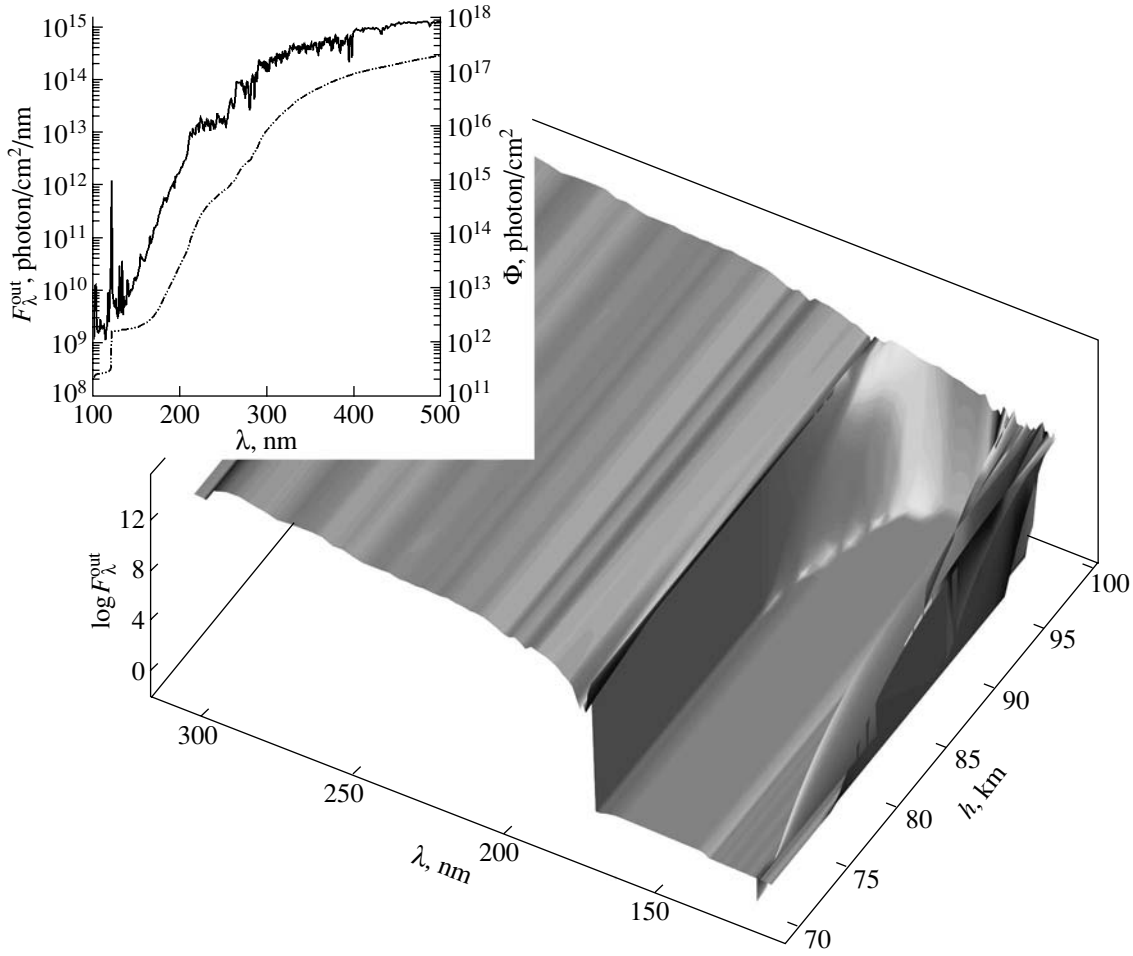


Fig. 2. Spectral power distribution of solar radiation versus altitude for the polar mesosphere under summer conditions. The zenith angle is 85°. Inset shows the extra-atmospheric flux F_λ^{out} (solid curve) and integral flux $\int_0^\lambda F_\lambda^{out} d\lambda$ (dash–dot–dot curve) of solar radiation.

The presence of dust grains affects not only charged species in the ionosphere, but also its neutral components. In particular, catalytic production of water molecules on microparticle surface can increase the water concentration in dusty regions, which promotes particle nucleation, i.e., giving rise to negative feedback. Note that the effect of dust grains on minor components of the middle atmosphere is poorly studied to this day. However, it is a promising area of research in photochemistry of the middle atmosphere, because the rates of heterogeneous chemical reactions on microparticle surface are comparable to those of gas-phase chemical reactions in the middle atmosphere. For example, there exists evidence of reduced concentration of atomic oxygen in the vicinity of NLC [17]. Note also that the formation of NLC and PMSE may be related to global climate change, which additionally motivates the study of these structures. It is well established that the frequency of appearance of PMC has considerably increased over the past decades [9], while the temperature distribution in the polar mesosphere under summer conditions has remained virtually the same[2] [18]. This trend can be attributed to increasing water vapor concentration in the mesosphere due to increasing methane concentration (whose oxidation is a source of water). It is well known that the methane concentration in the lower atmosphere also tends to increase, but the cause-and-effect relationships responsible for correlations between these trends have not been established.

One important characteristic of the polar mesosphere under summer conditions is its irradiation by solar radiation. Figure 2 shows the solar radiation spectra versus altitude computed for a zenith angle of 85°. In addition, we present the spectral power distribution

[2] The increase in the frequency of appearance of PMC may be explained by an additional (yet unrevealed) cooling of the mesosphere due to man-made emission of carbon dioxide into the lower atmosphere. According to current views, increase in greenhouse CO_2 concentration leads to temperature increase in the lower atmosphere and its decrease in the mesosphere.

F_λ^{out} of solar radiation and the corresponding energy flux

$$\Phi(\lambda) = \int_0^\lambda F_\lambda^{out} d\lambda.$$

The latter can be used to evaluate the role played by the photoelectric emission from dust-grain surface for microparticle materials having various optical and photoelectric properties. The solar radiation spectra were computed by using the PHODIS code [19].

The chosen zenith angle is the most common. In summer, the Sun hardly ever sets at the altitudes considered here, and the solar radiation spectrum weakly depends on the zenith angle α_s if $\alpha_s \le 90°$. Moreover, even if $\alpha_s \le 95°$, then scattered radiation with wavelengths $\lambda \ge 300$ nm that reaches the mesosphere is almost unattenuated (only relatively short-wavelength components of solar radiation are strongly absorbed). The zenith-angle range of 85–95° almost entirely covers the experimental data base on PMCs. It is clear that solar radiation with $\lambda \le 175$ nm (photon energy $\hbar\omega \ge 7.3$ eV) is strongly absorbed at altitudes $h \approx 100$ km. The work function W_{ice} for ice is about 8.7 eV. Therefore, there is no photoelectric emission from ice particles. In particular, this implies that ice microparticles are negatively charged. However, if a dust grain contains metal inclusions (which are characterized by a work function W of a few electronvolts, e.g. [20], W = 2.35 eV for sodium, W = 2.22 eV for potassium, W = 2.80 eV for calcium, W = 3.64 eV for magnesium, W = 4.2 eV for aluminum, and $W \approx 4.6$ eV for iron), then photoelectric emission can play a certain role, depending on the grain size, in the ionization balance in the middle atmosphere and the grain may be positively charged. This is possible for microparticles produced when micrometeorites containing substantial amounts of these metals are burned. The ionization rate q_d^{photo} corresponding to a single metal microparticle can be comparable to natural ionization sources, which include ionizing solar radiation and fast electrons:

$$q_d^{photo} \approx Q_{abs}\Phi(\lambda_W)Y(\lambda \le \lambda_W)$$
$$\sim 0.1\text{–}10\ \text{cm}^{-3}\ \text{s}^{-1} \sim q_e.$$

Here, $Q_{abs} \approx \pi a^2$ is the effective cross section for absorption of solar radiation by a metal microparticle of radius a, $\Phi(\lambda_W)$ is the integral flux of photons having energies higher than W, and $Y(\lambda)$ is the photoelectron yield (normally, $Y \sim 10^{-2}$–10^{-4}). Thus, the charge carried by the microparticles contained in the mesosphere may be of either sign, depending on their composition, which complicates their effect on the composition of the ionospheric plasma.

Despite the fact that the history of sounding rocket flights in the polar mesosphere under summer conditions spans more than thirty years, the data obtained are not sufficient to determine the plasma composition in dust structures. Both electron and ion concentrations were measured only in eleven experiments. Microparticle size distributions were measured only in two. Since the ionization conditions in the mesosphere depend on many factors, including ionization rate, neutral-gas composition, solar radiation flux, dust-grain composition, and neutral-gas temperature, the data acquired in the experiments can hardly be interpreted unambiguously. Nevertheless, a systematic analysis of experimental results concerning the behavior of charged species in the polar mesosphere under summer conditions shows that five trends can be identified [10, 21, 22]:

(i) considerable dips ("bite-outs") in both electron and positive-ion concentrations at altitudes between 80 and 85 km;

(ii) considerable decrease in the electron concentration without any significant change in the concentration or composition of the positive-ion subsystem;

(iii) a considerable increase in electron concentration in the region occupied by NLC while no measurements of ion concentrations were performed;

(iv) a considerable increase in the concentration of positive ions accompanied by a dip in electron concentration;

(v) no appreciable change in the electron and positive-ion concentrations while strong radio reflections were observed in the mesosphere (at altitudes between 80 and 85 km) in radar frequency bands of 50 to 1000 MHz.

It is shown below that these results can be explained by the presence of microparticles in the middle atmosphere.

3. THEORETICAL MODEL OF DUST STRUCTURES

The model of dust structures developed here relies on the data concerning the polar mesosphere under summer conditions summarized below. The unperturbed charged component is described by invoking a model of the polar ionosphere [23], which provides a quantitatively accurate description of the diurnal variation of the ionospheric plasma density at altitudes of 80 to 100 km. To simplify analysis, we consider two groups of positive ions: the simple ions N_2^+, O_2^+, and NO^+ are characterized by $\alpha_{rec}^s \approx 10^{-7}$ cm^3 s^{-1}; the proton hydrates $H^+(H_2O)_n$ are characterized by $\alpha_{rec}^c \approx 10^{-5}$ cm^3 s^{-1} averaged over n. Both the rate β_c of aggregation of simple ions into clusters and the ionization rate q_e are determined by using the aforementioned model of the polar ionosphere under summer conditions. Under the condi-

tions considered here, the typical values of β_c and q_e are $\beta_c \sim 0.1$ s^{-1} and $q_e \sim 0.1$–10 cm^{-3} s^{-1}.

The transfer of solar radiation in the middle atmosphere is computed by taking into account the actual value of zenith angle in a particular season. To describe the neutral composition, we use the model of the polar atmosphere under summer conditions developed in [24]. A model vertical distribution of microparticles is used as an initial one, with a peak concentration $n_d \sim 10^3$ cm^{-3} at an altitude of 90 km. The characteristic scale of variation of microparticle concentration is comparable to the altitude scale H of the middle atmosphere ($H \approx 7$ km). The typical neutral-gas concentration n_n at altitudes of about 80 km are determined by the model of the atmosphere: $n_n \approx 5 \times 10^{14}$ cm^{-3}.

The set of equations describing the effect of dust grains on the ionization characteristics of the ionosphere under summer conditions is written in a local approximation, since the characteristic charging time for microparticles in a dusty ionosphere is much shorter than those characterizing their transport. It includes balance equations [16] for the concentrations n_e, n_i^s, and n_i^c of electrons, simple ions, and clusters, respectively, and for the charge Z_d^a of a microparticles of radius a,

$$\frac{\partial n_e}{\partial t} = q_e - \alpha_{rec}^s n_e n_i^s - \alpha_{rec}^c n_e n_i^c + L_{photo}^e - L_{dust}^e, \quad (2)$$

$$\frac{\partial n_i^s}{\partial t} = q_e - \alpha_{rec}^s n_e n_i^s - \beta_c n_i^s - L_{dust}^s, \quad (3)$$

$$\frac{\partial n_i^c}{\partial t} = \beta_c n_i^s - \alpha_{rec}^c n_e n_i^c - L_{dust}^c, \quad (4)$$

$$\frac{\partial Z_d^a}{\partial t} = q_{photo} + \nu_s + \nu_c - \nu_e, \quad (5)$$

and a kinetic equation for the microparticle velocity distribution function $f_d(h, a, v, t)$ at an altitude h,

$$\frac{\partial f_d}{\partial t} + \frac{\alpha_w m_w v_w^{th} \rho_d [H_2O]}{4}\frac{\partial f_d}{\partial a} + v\frac{\partial f_d}{\partial h} + \left(g - \frac{\pi \rho c_s a^2 F_d (v + v_{wind})}{m_d}\right)\frac{\partial f_d}{\partial v} = 0. \quad (6)$$

Here, L_{dust}^j ($j = e$, s, c) represents the losses of electrons, simple ions, and clusters in collisions with dust grains,

$$L_{dust}^j = \int \nu_j dn_d$$

(the microparticle size distribution, $n_d = \int f_d dv da$, is taken into account in dn_d); the term L_{photo}^e represents the photoelectric emission induced by solar radiation,

$$L_{photo}^e = \int q_{photo} dn_d,$$

and the terms ν_e, ν_s, and ν_c represent the charging rates due to collisions of electrons, simple ions, and clusters with microparticles, respectively. The term q_{photo} represents the ionization rate due to photoelectric emission; β_c, the rate of aggregation of simple ions into clusters.

In Eq. (6), the second and fourth terms describe, respectively, microparticle growth in the ambient supersaturated water vapor and either sedimentation or rise of dust grains subject to neutral drag. Here, ρ and ρ_d denote the densities of the ambient air and grain material, respectively; m_d is the grain mass; α_w is the accommodation coefficient for water molecules colliding with a dust grain (normally, $\alpha_w \sim 1$); v_w^{th} is the thermal velocity of water molecules; c_s is the local speed of sound; v_{wind} and v are the vertical wind and grain velocities, respectively; and the factor F_d (on the order of unity) reflects the effect of grain geometry.

It should be noted that the large-scale circumpolar vortex that forms in the Earth's atmosphere under the conditions considered here, in particular, induces upward motion of the gas entrained in the vortex motion. In the mesosphere, this gives rise to an upward component of the mean neutral-gas velocity, $v_{wind} \approx$ 1–10 cm/s.

The coagulation of colliding microparticles in the mesosphere can be neglected, since the corresponding time scale is much greater than other characteristic times:

$$\tau_{coag} \approx (n_d v \pi a^2)^{-1} \geq 10^6 \text{ s}.$$

Equations (2)–(6) describe the evolution of a particle located at an altitude h. This approximation is applicable because the vertical PMC size (~1 km) is much smaller than their horizontal extent (~100 km), while the horizontal velocity characteristic of microparticle transport is less than, or comparable to, their vertical velocity, and the horizontal particle displacement over times on the order of 24 hours is negligible.

Microscopic electron and ion currents incident on the grains are calculated by using the orbit-limited probe model [25, 26], in which the cross sections for collisions of electrons and ions with a charged grain are determined by angular-momentum and energy conservation laws. For a negatively charged grain, the orbit-

limited probe approximation leads to the following expressions for charging rates:

$$\nu_e \approx \pi a^2 \left(\frac{8T_e}{\pi m_e}\right)^{1/2} n_e \exp\left(\frac{eq_d}{aT_e}\right),$$

$$\nu_i \approx \pi a^2 \left(\frac{8T_i}{\pi m_i}\right)^{1/2} n_i \left(1 - \frac{eq_d}{aT_i}\right).$$

For a positively charged grain, ν_e and ν_i are expressed as

$$\nu_e \approx \pi a^2 \left(\frac{8T_e}{\pi m_e}\right)^{1/2} n_e \left(1 + \frac{eq_d}{aT_e}\right),$$

$$\nu_i \approx \pi a^2 \left(\frac{8T_i}{\pi m_i}\right)^{1/2} n_i \exp\left(-\frac{eq_d}{aT_i}\right).$$

Here, $q_d = Z_d e$ is the grain charge; $-e$ is the electron charge; T_e and T_i are the electron and ion temperatures, respectively; and m_α is the particle mass for the specie denoted by α. The ionization rate q_{photo} associated with photoelectric emission is determined by the solar radiation flux $F(\lambda)$:

$$q_{photo} \approx \pi a^2 \int_0^{\lambda^*} Q_{abs}(a, m(\lambda)) F(\lambda) Y(\lambda) d\lambda, \qquad (7)$$

where Q_{abs} is the photon absorption cross section,

$$m(\lambda) = n(\lambda) + i\kappa(\lambda)$$

is the complex refractive index of the grain material, $F(\lambda)$ is the solar radiation energy flux, λ^* is the longest wavelength of a photon inducing photoelectric emission, and $Y(\lambda)$ is the photoelectric emission probability. The last quantity is usually calculated by using the Fowler–Nordheim relation [27]

$$Y(\lambda) = C\left(W - \frac{2\pi\hbar c}{\lambda}\right)^2,$$

where the factor C varies within 10^{-2}–10^{-4} eV^{-2}, depending on the grain size and composition [28]. Note that photoelectric emission plays a significant role when the grain size is sufficiently large. In the case of a particle size of a few nanometers, the photoelectric emission is weak, since the absorption cross section Q_{abs} decreases with a as a^3 when $a \ll \lambda$.

The integration limits in (7) are set by the condition that only photons with energies such that

$$\frac{2\pi\hbar c}{\lambda} - W - \frac{e^2 Z_d}{a}$$

create photoelectrons. In the case under analysis, the photon energy is a few electronvolts, while the equilibrium potential of a charged grain is on the order of

$$\frac{T_e}{e} \approx \frac{T_n}{e},$$

where $T_n \approx 0.03$ eV is the temperature of neutrals in the mesosphere. Indeed, since the initial energy of a photoelectron is a few electronvolts,

$$E \approx \frac{2\pi\hbar c}{\lambda} - W \sim 1 \text{ eV} \gg T_n,$$

the charging of dust grains depends on their thermalization rate. The photoelectron temperature can be estimated using the equation

$$\frac{\partial T_e}{\partial t} = -\nu_{en}(T_e)\delta_{en}(T_e)(T_e - T_n), \qquad (8)$$

where $\nu_{en}(T_e)$ is the electron–neutral collision frequency and $\delta_{en}(T_e)$ is the fraction of electron energy transferred in a collision. When $T_e \sim 1$ eV, electrons cool down most effectively via inelastic collisions with molecules of O_2 or N_2 (since vibrational states of the molecules are excited by electron impact), in which case $\delta_{en} \approx 10^{-2}$. The corresponding electron cooling time at mesospheric altitudes is

$$\tau_c^2 \sim (\nu_{en}\delta_{en})^{-1} \sim 10^{-2} \text{ s},$$

i.e., much shorter than the dust-grain charging time ν_i^{-1} estimated as

$$\nu_i^{-1} \sim (\pi a^2 n_i v_{Ti})^{-1} \geq 0.1 \text{ s}.$$

Therefore,

$$T_e \approx T_n$$

and, accordingly,

$$\frac{e^2|Z_d|}{a} \sim kT_e \approx 0.01 \text{ eV} \ll \frac{2\pi\hbar c}{\lambda}.$$

Thus, photoelectric current from a dust grain can be neglected. Recall also that the effect of negative ions on microparticle-charging processes in the mesosphere is neglected here.

The set of Eqs. (2)–(7) provides a self-consistent description of the spatiotemporal variation of ionization characteristics (plasma density, ion composition, and microparticle charge) and the microparticle size distribution in the polar mesosphere under summer conditions. Boundary conditions for the microparticle

velocity distribution function $f_d(t, a, h)$ are set as follows: at the upper boundary (h = 100 km),

$$f_d(a, h, t) = 0,$$

which corresponds to the absence of microparticles at high altitudes; at the lower boundary (h = 78 km), we also set

$$f_d(a, h, t) = 0.$$

In physical terms, this means that microparticles rapidly evaporate as they leave the region occupied by supersaturated water vapor.

4. NUMERICAL RESULTS

In this section, we discuss the results obtained by applying the theoretical model presented above to the polar ionosphere at 80 to 95 km under summer conditions. Our discussion is focused on the possibility of explaining the most important experimental data concerning behavior of charged species in a dusty ionosphere.

Figure 3 shows the mean equilibrium charge of a dust grain of radius 100 nm located at an altitude of 85 km as a function of the microparticle concentration and the effective solar radiation flux that induces photoelectric emission. This parametric study facilitates estimation of the dependence of microparticle charge on the zenith angle under summer conditions in the

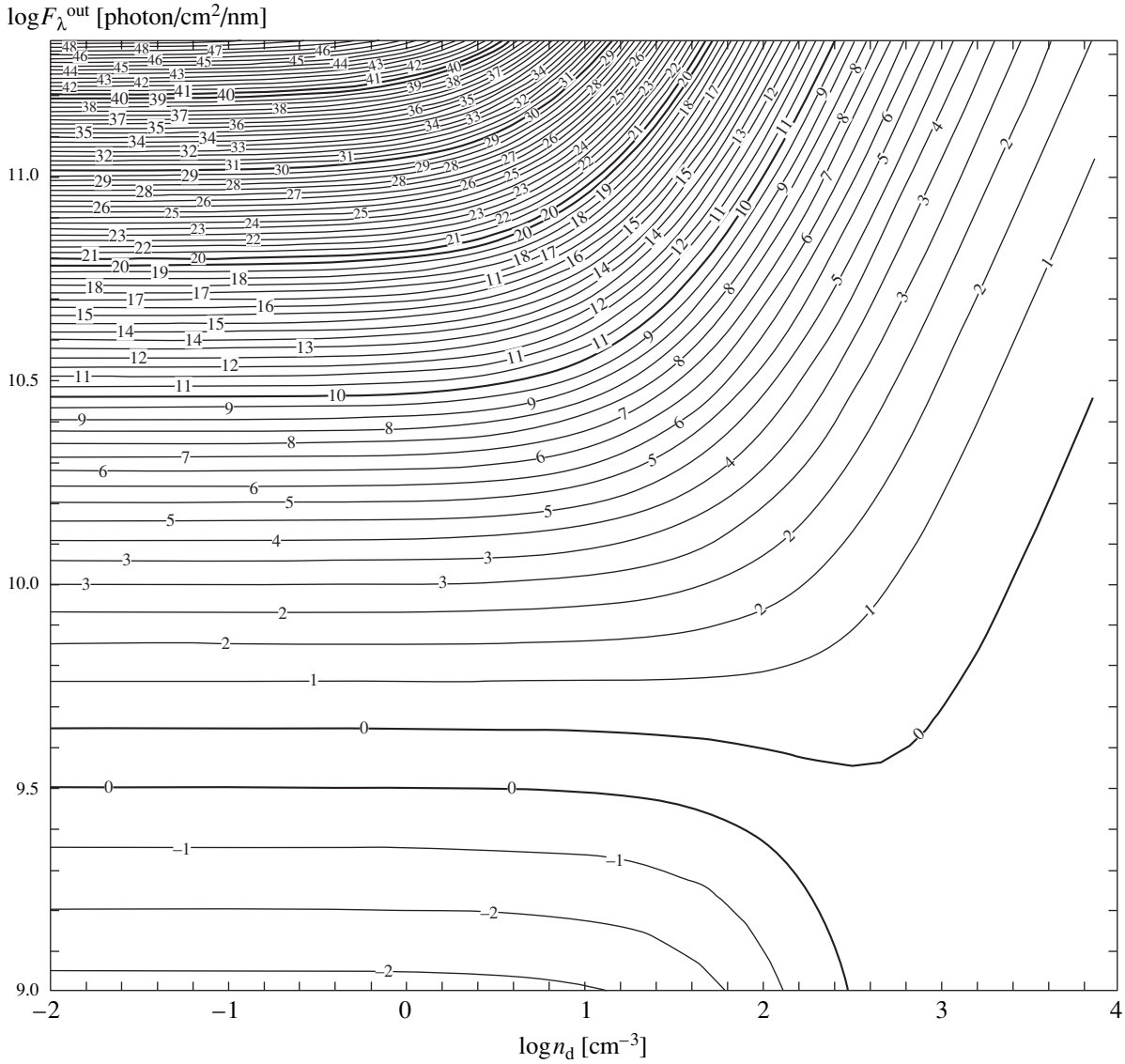


Fig. 3. Mean charge of a dust grain of diameter 100 nm at an altitude of 85 km (shown at curves in units of electron charge) versus solar radiation energy flux and dust-grain concentration. The work function for the grain material is about 4 eV.

polar mesosphere. The work function is assumed to be about 4 eV; i.e., an ice particle containing metal inclusions is considered. Figure 3 demonstrates that, even when exposed to a moderate solar radiation flux, a microparticle carries a relatively large positive charge $Z_d \approx 50$, which decreases with increasing grain concentration. For typical dust-grain concentrations in the mesosphere ($n_d \approx 300$ cm^{-3}) and moderate solar radiation flux, the mean dust-grain charge is about ten electron charges. It is clear that the plasma composition in the region occupied by dust grains changes, because the total charge carried by the microparticles is comparable to that of electrons and ions:

$$n_d|Z_d e| \sim |e|n_e.$$

As the effective solar radiation flux decreases, the mean dust-grain charge becomes negative and comparable to the electron charge. This implies that the absolute charge carried by a grain in the mesosphere can be substantially reduced by photoelectric emission. (Note that the dust-grain charge in the absence of photoelectric emission is negative, and its mean absolute value is about ten electron charges.)

Under the conditioned under consideration, charge fluctuations are relatively large,

$$|\delta Z_d| \sim |Z_d|,$$

and Eq. (5) cannot be used to describe the charging of a dust grain, because it is derived under the implicit assumption that the charge varies continuously rather than by discrete portions. When Z_d is small, the evolution of the mean dust-grain charge is described by the Langevin equation [29]

$$\frac{\partial Z_{d,j}}{\partial t} = q_{\text{photo},j} + \nu_{i,j}^{p} + \nu_{i,j}^{c} - \nu_{e,j} + g(t),$$

where $g(t)$ is a zero-mean random function:

$$g(t)g(t+t') \approx 2\tau_{ch}^{-1}(\delta Z_d)^2\delta(t-t'),$$

with $\tau_{ch} \sim \nu_i^{-1}$ denoting the characteristic dust-grain charging time. The difference between the continuous and discrete charging equations can be important for estimating the mean charge of a relatively small grain (about 10 nm in diameter), which appear to be of minor importance for the physics of polar mesospheric clouds. For dust grains of interest here (having diameters greater than 10 nm), the mean charge is relatively large ($|Z_d| \gg 1$), and Eq. (5) is applicable.

Figure 4 shows the evolution of vertical grain size distribution and the mean grain size versus time and altitude. The initial vertical profile is the Gaussian distribution of 10-nm grains

$$n_d \propto \exp(-(h-h_0)^2/H^2),$$

where $h_0 = 90$ km and $H \approx 7.5$ km. The solution to Eq. (6) presented here illustrates the formation of a layer of dust grains of characteristic diameter $2a \approx 0.5$ μm at altitudes of about 80 km over a few hours, with a grain concentration of $n_d \approx 10$–1000 cm^{-3}). It should be noted that the characteristic diameter depends both on the accommodation coefficient α_w and on the density ρ_d of a growing particle. The figure clearly demonstrates the growth of sedimenting particles. The grain residence time in the mesosphere increases in the presence of upward air flow, and the characteristic grain diameter increases accordingly. The inset illustrates the effect of upward air flow on the characteristics of a single grain of radius $r_0 = 3$ nm initially located at $h_0 = 85$ km. The vertical location h_d and radius r_d of the grain are plotted versus time. The vertical wind speed v_{wind} was set equal to 10 cm/s, which is close in order of magnitude to wind speeds observed in the polar mesosphere under summer conditions. With time elapsed, $r_d \longrightarrow r_\infty \approx 200$ nm and $h_d \longrightarrow h_\infty = 80$ km. The values of r_∞ and h_∞ are independent of h_0 and r_0, being determined only by v_{wind}. Therefore, when the upward velocity of air motion in the mesosphere is not zero, the grains leaving a dust cloud rapidly evaporate, their size decreases, and upward air motion brings them back into the mesosphere, where they grow and settle again. As a result, a sharp lower boundary of the dust cloud appears at the altitude h_b determined by the condition

$$\Phi_{H_2O}(h_b) = \frac{P_{H_2O}(h_b)}{P_{\text{sat}}(h_b)} = 1.$$

For a grain of radius $r_0 = 3$ nm, the minimal upward velocity required to bring a grain back into the mesosphere is relatively low,

$$v_{\text{wind}} \approx \frac{\rho_d}{\rho}\frac{g}{c_s}r_0 \approx 0.1 \text{ cm/s},$$

and can easily be reached under summer conditions in the polar mesosphere by virtue of the existence of the circumpolar vortex. Thus, the sedimentation and growth of nanometer-scale particles in the mesosphere lead to the development of a narrow layer of submicron-sized particles at altitudes of 80 to 85 km, i.e., where noctilucent clouds are observed.

The presence of submicron-sized particles in the mesosphere has a strong effect on plasma characteristics (depending on photoelectric properties of dust-grain material). This effect is illustrated in Fig. 5 by

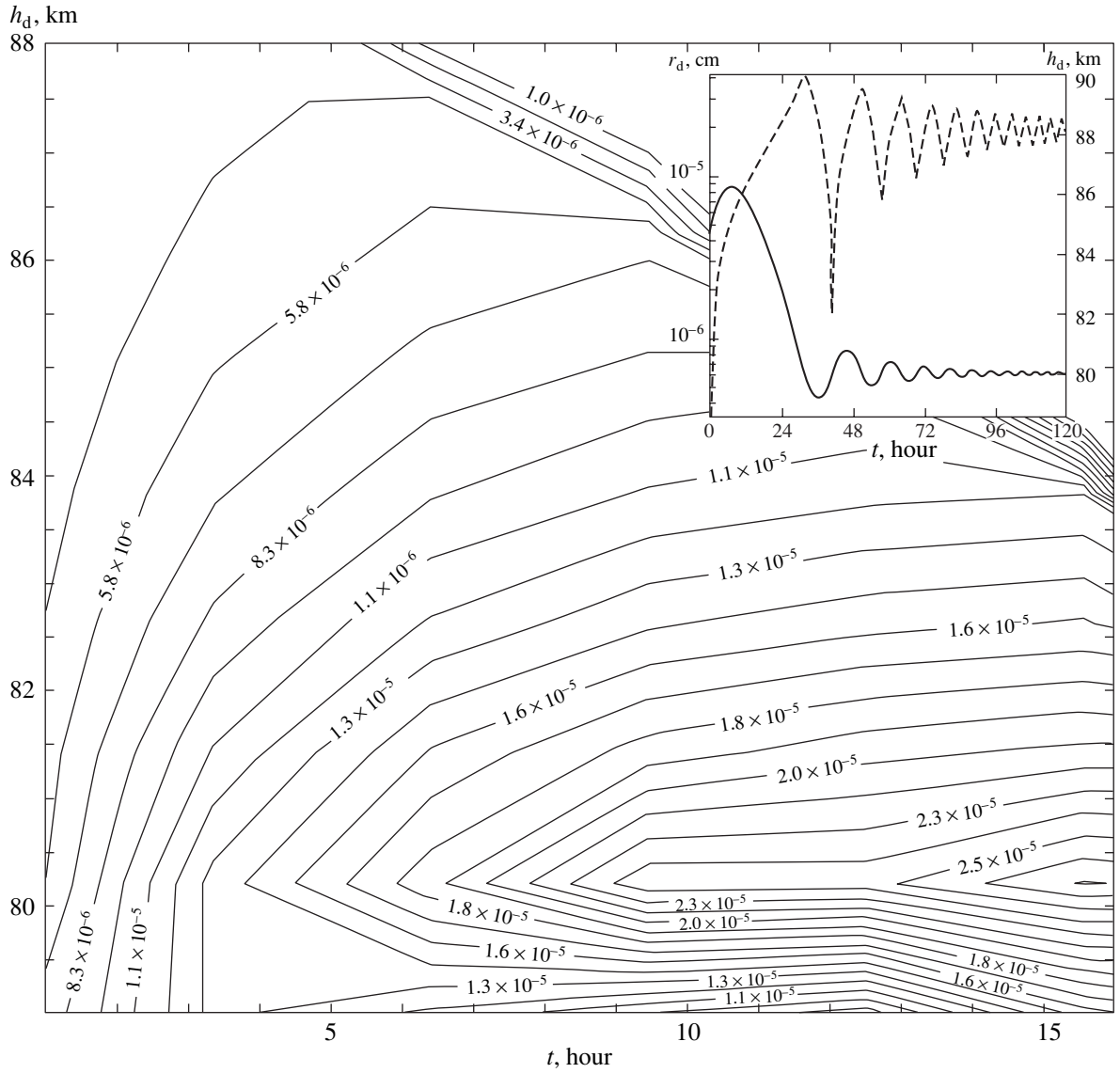


Fig. 4. Mean dust-grain size (shown at curves in centimeters) versus time and altitude as a characteristic of dust sedimentation in the polar mesosphere under summer conditions. Formation of a layer of particles 0.3–0.5 μm in diameter at altitudes of about 80 km over a time period of about 10 h is demonstrated. The inset shows typical altitude (solid curve) and radius (dashed curve) of a dust grain versus time as characteristics of the effect of upward air flow on particle motion. Vertical air velocity is 10 cm/s.

spatiotemporal variations of mesospheric ion and electron concentrations in the presence of a layer of dust grains characterized by different values of the work function. The right panels correspond to pure ice microparticles. In this case, since there is no photoelectric emission, the microparticles are negatively charged and both ion and electron concentrations are reduced in the region occupied by the dust layer. Indeed, the presence of dust grains in the ionosphere leads to recombination of the plasma on the grain surface. As a result, the rates of electron and ion losses increase according to Eqs. (2)–(5), and the corresponding concentrations decrease. It can be shown that a concentration $n_d \approx 10^2$ cm^{-3} of dust grains of diameter $2a \approx 100$ nm is sufficient to substantially reduce the mesospheric plasma density. These dust-grain parameters are common for NLC-type structures. It should be noted that a decrease in electron concentration reduces the loss of ions due to recombination. Therefore, while the presence of microparticles reduces the electron concentration, the ion concentration in the mesosphere may remain unchanged or even increase. These effects obviously depend on the relative importance of the ion losses due to recombination and collisions with dust grains. As the dust-grain concentration exceeds a certain value, the equilibrium ion concentration becomes independent of the electron concentration and is controlled only by the loss of ions in collisions with microparticles, because the electron concentration is substantially reduced and

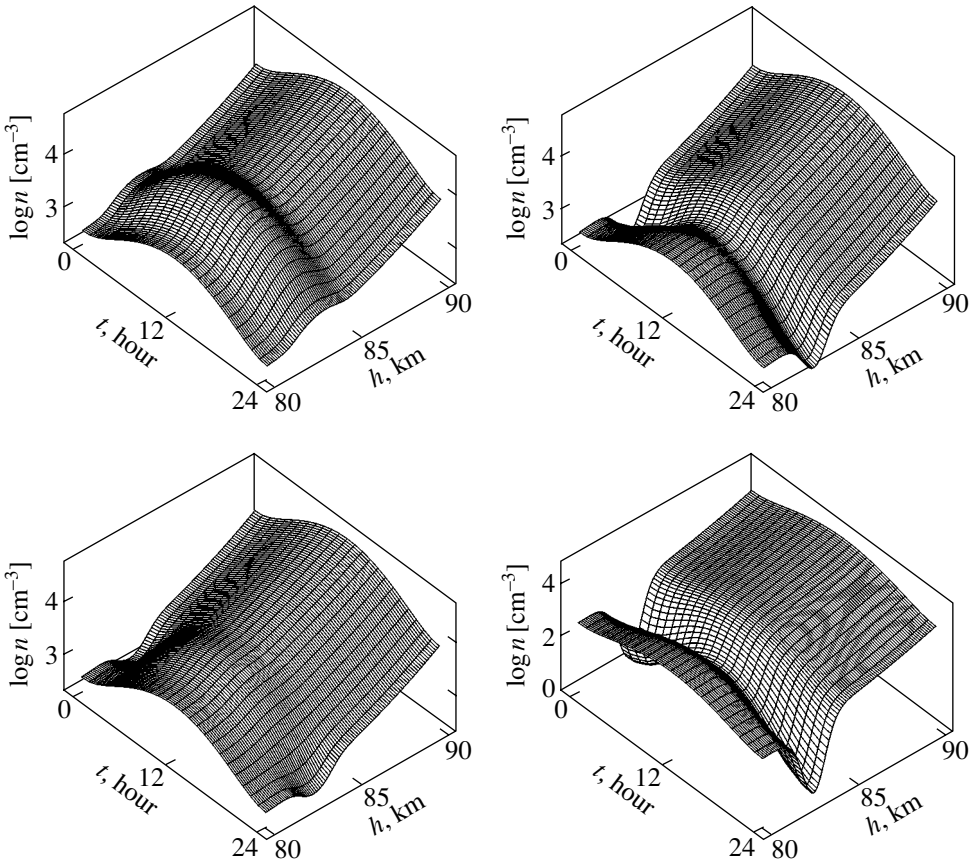


Fig. 5. Effect of the photoelectric properties of a dust layer on the electron and ion concentrations in the lower ionosphere. The right panels correspond to the absence of photoelectric emission from dust grains (the work function exceeds 7.3 eV, as in the case of pure ice microparticles). The left panels illustrate strong effect of photoelectric emission on the plasma composition in the mesosphere (work function below 7.3 eV, as in the case of calcium microparticles). Top and bottom panels are electron and ion concentrations, respectively.

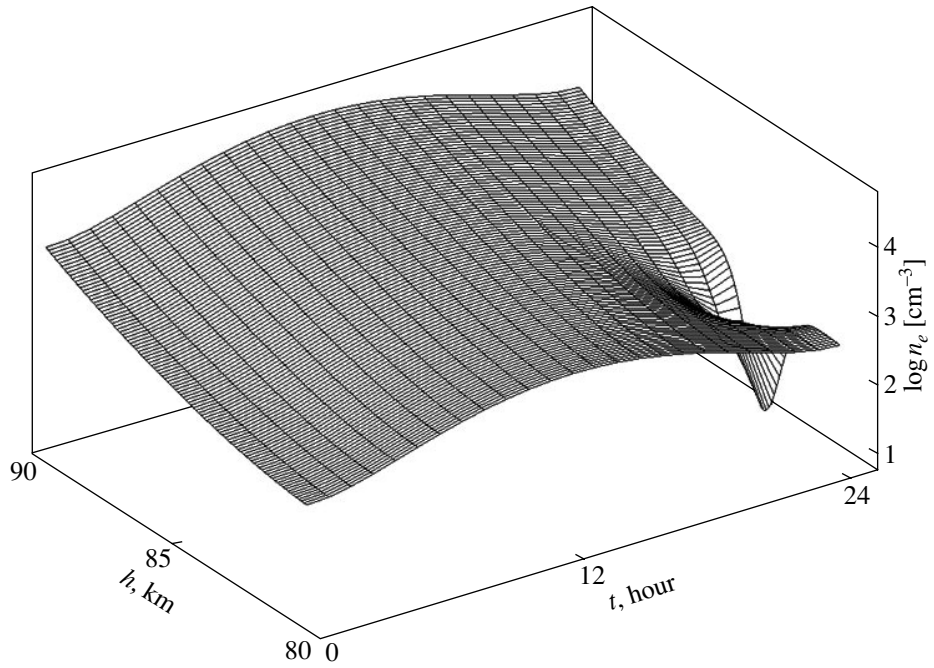


Fig. 6. Electron concentration in a dusty mesosphere versus altitude and time under summer conditions. Initially, dust grains are a few nanometers in diameter and their presence has no effect on n_e. Their growth and sedimentation lead to the formation of a thin layer of charged submicron-sized particles and a lower electron concentration at altitudes of 80 to 85 km.

the role played by ion losses due to recombination is insignificant.

The left panels in Fig. 5 correspond to a microparticle material characterized by a relatively low work function ($W \approx 3$ to 4 eV). In this case, even if the microparticle concentration is low ($n_d \sim 10$–100 cm^{-3}), the present model predicts an increase in electron concentration and a decrease in ion concentration, due to higher recombination losses.

Note that the mesospheric plasma exhibits a more diverse behavior when photoelectric emission plays a significant role. Depending on the microparticle size and concentration, virtually any effect due to the presence of microparticles in the mesosphere (see Section 2) can be observed.

Figure 6 shows the evolution of electron concentration in the mesosphere associated with sedimentation and growth of dust grains. The plotted solution to Eqs. (2)–(7) illustrates the self-consistent behavior of the electron plasma component under summer conditions in the polar mesosphere (correlated with dust-grain growth and solar radiation flux). It is clearly demonstrated that the microparticle layer begins to substantially modify the state of the ionospheric plasma only after several hours have passed, i.e., when the grain size has substantially increased. The formation of such a layer is characterized by a time scale of about 24 hours. This explains the experimental fact that neither NLC nor PMSE can be observed permanently (while the water vapor contained in the mesosphere is supercooled). Indeed, if the characteristic time $\tau_{\text{profile}}^{\text{dust}}$ required for nanometer-sized particles to fill the mesosphere is long as compared to the time of dust-layer formation (about 24 h in the case considered here), then the frequency of appearance of mesospheric dust clouds (NLC and PMSE) is controlled only by $\tau_{\text{profile}}^{\text{dust}}$. According to observations of noctilucent clouds, this time scale substantially exceeds the characteristic time of dust-layer formation.

5. CONCLUSIONS

We have developed a theoretical model that provides a self-consistent description of spatiotemporal variations of the plasma composition in the polar mesosphere under summer conditions and formation of noctilucent clouds and polar mesospheric summer echoes. The model describes sedimentation of dust grains in the middle atmosphere, their growth in a supersaturated water vapor, and microparticle charging processes, allowing for variations of the ion-subsystem composition in the polar mesosphere and photoelectric emission. The model can be used to predict the five trends associated with NLC and PMSE enumerated above: considerable increase or decrease in the electron concentration observed at altitudes between 80 and 85 km, with or without significant change in the concentration or composition of the positive-ion subsystem. The model provides a unified explanation of these trends and of the presence of both positively and negatively charged particles in the upper atmosphere discovered by means of sounding rocket flights, as well as of the experimental fact that neither NLC nor PMSE can be observed permanently.

ACKNOWLEDGMENTS

This work was conducted under the basic research program “Nanoparticles in Nature: Localization Conditions, Technological and Ecological Consequences” supervised by the Division of Geosciences of the Russian Academy of Sciences and was supported, in part, by INTAS, project nos. 97-2149 and 01-0391 and by the Russian Foundation for Basic Research, project no. 03-05-64813a. We thank the referee for careful study of the manuscript and valuable recommendations that helped to improve presentation.

REFERENCES

1. *Dusty Plasmas*, Ed. by A. Bouchoule (Wiley, Chichester, 1999).
2. J. H. Chu and I. Lin, Phys. Rev. Lett. **72**, 4009 (1994).
3. H. Thomas and G. E. Morfill, Nature **379**, 806 (1996).
4. V. N. Tsytovich, Usp. Fiz. Nauk **167**, 57 (1997) [Phys. Usp. **40**, 53 (1997)].
5. S. I. Popel, A. P. Golub’, and T. V. Loseva, Pis’ma Zh. Éksp. Teor. Fiz. **74**, 396 (2001) [JETP Lett. **74**, 362 (2001)].
6. V. E. Fortov, A. G. Khrapak, S. A. Khrapak, *et al.*, Usp. Fiz. Nauk **174**, 495 (2004) [Phys. Usp. **47**, 447 (2004)].
7. R. P. Turco, O. B. Toon, R. C. Whitten, *et al.*, Planet. Space Sci. **30**, 1147 (1982).
8. M. Gadsden and W. Schröder, *Noctilucent Clouds* (Springer, Berlin, 1989).
9. J. D. Mathews, J. Atmos. Sol.–Terr. Phys. **60**, 413 (1998).
10. J. Y. N. Cho and J. Röttger, J. Geophys. Res. **102**, 2001 (1997).
11. O. Havnes, J. Trøim, T. Blix, *et al.*, J. Geophys. Res. **101**, 10829 (1996).
12. *Hazards Due to Comets and Asteroids*, Ed. by T. Gehrels (Univ. of Arizona Press, Tucson, 1994).
13. R. A. Goldberg, R. F. Pfaff, R. H. Holzworth, *et al.*, Geophys. Res. Lett. **28**, 1407 (2001).
14. G. C. Reid, J. Atmos. Sci. **25**, 523 (1975).
15. E. Kopp, Adv. Space Res. **25**, 173 (2000).
16. B. A. Klumov, S. I. Popel, and R. Bingham, Pis’ma Zh. Éksp. Teor. Fiz. **72**, 524 (2000) [JETP Lett. **72**, 364 (2000)].
17. B. J. Murray and J. M. C. Plane, Adv. Space Res. **31**, 2075 (2003).
18. F. J. Lübken, Geophys. Res. Lett. **27**, 3603 (2000).
19. A. Kylling, http://kaja.gi.alaska.edu.

20. A. G. Khrapak and I. T. Yakubov, *Electrons in Dense Gases and Plasmas* (Nauka, Moscow, 1981) [in Russian].
21. M. Rapp and F. J. Lübken, J. Atmos. Sol.–Terr. Phys. **63**, 759 (2001).
22. F. Balsiger, E. Kopp, M. Friedrich, *et al.*, Geophys. Res. Lett. **23**, 93 (1996).
23. A. D. Danilov, N. V. Smirnova, T. A. Blix, and E. V. Thrane, Adv. Space Res. **29**, 6 (2002).
24. G. P. Anderson, S. A. Clough, F. X. Kneizys, *et al.*, Report A371571 (Air Force Geophysics Lab., Hansom, MA, 1986).
25. F. F. Chen, in *Plasma Diagnostic Techniques*, Ed. by R. H. Huddlestone and S. L. Leonard (Academic, New York, 1965), Chap. 4.
26. M. S. Barnes, J. H. Keller, J. C. Forster, *et al.*, Phys. Rev. Lett. **68**, 313 (1992).
27. B. Feuerbacher and B. Fitton, J. Appl. Phys. **43**, 1563 (1972).
28. A. Schmitt-Ott, P. Schurtenberger, and H. C. Siegmann, Phys. Rev. Lett. **45**, 1284 (1980).
29. G. E. Morfill, V. N. Tsytovich, and H. Thomas, Fiz. Plazmy (Moscow) **29**, 3 (2003) [Plasma Phys. Rep. **29**, 1 (2003)].
30. V. Nussbaumer, K.-H. Fricke, M. Langer, *et al.*, J. Geophys. Res. **101**, 19161 (1996).
31. B. Smiley, S. Robertson, M. Horányi, *et al.*, J. Geophys. Res. **108**, 8444 (2003).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 100, No. 1, 2005, pp. 165–198.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 1, 2005, pp. 186–219.

STATISTICAL, NONLINEAR, AND SOFT MATTER PHYSICS

A Converging Equation of State of a Weakly Nonideal Hydrogen Plasma without Mystery

A. N. Starostin* and V. C. Roerich

Troitsk Institute for Innovation and Fusion Research (Russian State Scientific Center), Troitsk, Moscow oblast, 142190 Russia

**e-mail: staran@triniti.ru*

Received June 18, 2004

Abstract—A detailed independent derivation of the equation of state of a weakly nonideal hydrogen plasma is presented. The impetus for this work was the demand for high accuracy of the equation of state of the solar plasma in relation to the problems of modern helioseismology, accuracy sufficient for reproducing the velocity of sound on the Sun from optical measurement results with errors not exceeding 10^{-4}. The existing equations for the second virial coefficient in the expansion of the Helmholtz thermodynamic potential for a system of electrons and protons in powers of the activities of these particles involve certain procedures for the removal of the arising divergences that provoke questions and require independent verification. The suggested equation of state is used to qualitatively estimate the accuracy of various physical and chemical models. The speed of sound and adiabatic exponent calculated along the solar trajectory are presented for a model hydrogen plasma. The calculations were performed with the inclusion of relativistic corrections, electron degeneracy effects, radiation pressure in the plasma, Coulomb interaction in the Debye–Hückel approximation with diffraction and exchange corrections, and converging contributions of bound and scattering states. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Helioseismology opens up the unique possibility of very accurately checking the equation of state of a weakly nonideal plasma. It enables us to reproduce the local velocity of sound on the Sun with an accuracy higher than 10^{-4} from optical measurement results [1–3]. A comparison of various theoretical models with experiment can be used to check the existing methods for including the contributions of bound states and continuum represented within the framework of a physical model in the equations for the second virial coefficient [4–6].

It is commonly accepted, at least in the physical literature, that, starting with works [7–12] (these results can also be found in monographs [13, 14], which contain a comprehensive bibliography), the problem of the form of the equation of state of weakly nonideal hydrogen has been solved in principle, with an accuracy of Γ_D^2, where

$$\Gamma_D = e^2\kappa_D/T$$

is the Debye nonideality parameter (e is the charge of the electron, κ_D is the inverse Debye radius, and T is the temperature of the plasma in energy units). It is, in particular, assumed that the contribution of bound states (atomic partition functions) to the pressure of a plasma is given by the converging Planck–Brillouin–Larkin or, simply, Planck–Larkin (P–L) equation [6, 8–14].

At the same time, the atomic partition function is given in the astrophysical literature in the form suggested in [4], and the Planck–Larkin equation is open to criticism [15], because convergence in this equation arises starting with states with the principal quantum numbers

$$n \lesssim n_{\max} \sim \sqrt{\mathrm{Ry}/T},$$

where

$$\mathrm{Ry} = \frac{\mu e^4}{2\hbar^2} \approx 13.598 \text{ eV}$$

is the ionization potential of the hydrogen atom. Here,

$$\mu = \frac{m_e m_p}{m_e + m_p}$$

is the reduced mass (m_e and m_p are the masses of the electron and proton, respectively) and $\hbar$ is the Planck constant. However in reality, radiation from the solar photosphere contains Balmer series contributions with $n \approx 17 > 6 \gtrsim n_{\max}$ [16] at $T \approx 5800$ K. Irrespective of theoretical justification of various equations for the contribution of bound states to the equation of state for a hydrogen plasma in the innards of the Sun, helioseismology enables us to select those that better correspond to experiment. It turned out that the theory that uses the

Planck–Larkin equation is closer to experiment [6] than that described in [4], where the contribution of excited states is limited by ionization under the action of plasma microfields (the Unsöld mechanisms [17]).

The contribution of bound states was described in [18] by an equation different from the Planck–Larkin formula. More recently, this equation was substantiated in [19, 20]. In this work, attention is focused on the internally contradictory method for obtaining converging equations for the second virial coefficient that can be found in the literature (e.g., see [13, 14]) and that are based on Beth–Uhlenbeck-type equations [21], the Levinson theorem [22], and the methods for regularizing diverging equations [8, 10–12], whose applicability to long-range Coulomb interaction is not obvious.

Not infrequently, finite results are obtained in theoretical physics from initially inapplicable diverging equations, but, in certain instances (including the derivation of the equation of state of a weakly nonideal plasma), recipes used for this purpose contain elements of a certain mysticism and, if used uncritically, can lead to errors and contradictions to experiment.

The divergence of the atomic "partition function" Σ, including the summation over discrete and scattering states, is "removed" in [8] by subtracting contributions with one and two wavy lines that correspond to interaction in first-order (this contribution disappears because of electroneutrality) and second-order perturbation theory (this contribution has already been taken into account in the summation of a series of ring diagrams that yield a finite expression for Coulomb interaction in the Debye–Hückel approximation). The result given in [8] corresponds to regularization of the type

$$\Sigma_{\text{P–L}} \equiv \Sigma_{\text{reg}}(I) = \Sigma(I) - \Sigma(0) - I\left(\frac{\partial \Sigma}{\partial I}\right)_0, \qquad (1)$$

where I is the ionization potential of the hydrogen atom $I = \text{Ry}$,

$$\Sigma(I) = \sum_{n=1}^{\infty} n^2 \exp\left(\frac{I}{n^2 T}\right), \qquad (2)$$

$$\Sigma_{\text{P–L}} = \sum_{n=1}^{\infty} n^2 \left(\exp\left(\frac{I}{n^2 T}\right) - 1 - \frac{I}{n^2 T}\right). \qquad (3)$$

This regularization method for calculating the second virial coefficient was extended in [10] even to the subtraction of contributions up to the third-order derivative with respect to e^2 in the Taylor expansion of the integral in wavevectors that characterize scattering states in the Beth–Uhlenbeck equation.

In [11–13], the removal of diverging terms from the second virial coefficient was related to Debye screening in a plasma, which allowed the authors to apply the Beth–Uhlenbeck equation and the Levinson theorem.

We show in what follows that the divergences that arise when we calculate the contribution of bound states are only removed by a fragment of the diagram with one wavy line that corresponds to the inclusion of discrete states in the theorem of completeness of Coulomb wavefunctions that describe the relative motion of the electron–proton system. The inclusion of continuum states with the use of the theorem of completeness of Coulomb functions for this diagram together with the subtraction of the contribution of two crosspieces (already taken into account in obtaining the Debye–Hückel equation) removes divergences in the equation for the contribution of scattering states to the second virial coefficient.

The resulting converging equation for the atomic partition function is different from the Planck–Larkin equation. We also calculated the finite contribution of scattering states from all pair interactions to the second virial coefficient. These calculations required the search for correct methods for handling the squares of the wavefunctions of the continuous spectrum containing δ functions in the momentum representation. The equations for attractive states in the second virial coefficients were generalized to degenerate electrons, which appears to be important in view of astrophysical applications, for instance, for helioseismology, because the solar plasma is weakly nonideal, and the expansion of the equation of state up to the second virial coefficient taking into account electron degeneracy is quite justified.

Interestingly, the contribution of attractive states to the second virial coefficient summed over discrete and continuum states coincides with the equation given in [13] to within certain corrections in the logarithmic term. We will show that self-consistent calculations of the contributions of discrete and continuous spectra admit the use of the Planck–Larkin equation together with the properly determined contribution of scattering states. The problem of "correctly" taking into account the regularized atomic partition function arises within the framework of the chemical model, in which atoms (molecules) are declared "good" quasi-particles (found from calculations or experiment), whereas the contributions of interaction in the continuous spectrum are ignored in calculating the equation of state.

This approximation often becomes inevitable because of the complexity of describing an equilibrium multicomponent plasma, when the inclusion of scattering states requires a description of many-particle interaction (starting with HeI, for which we must solve the quantum scattering problem for three bodies, etc.).

We compare the complete physical model with chemical models that use various methods for regularizing the atomic partition function. For a weakly ionized plasma, the difference between these methods is inessential (if $I/T \gg 1$). For a hot plasma, such as the plasma close to the center of the Sun ($I/T \ll 1$), taking into account the discrete spectrum only, ignoring the

contribution of scattering states, leads to superfluous accuracy.

This work is organized as follows. In Section 2, we analyze the perturbation theory series and formulate a consistent method for taking into account plasma neutrality in the expansion in powers of activity. This is already necessary in calculations of exchange coupling contributions and the Debye–Hückel approximation with corrections for diffraction and the screening of "dressed" particles.

Section 3 analyzes the contribution of ladder diagrams to the equation for the second virial coefficient and formulates the regularization rules for the integration of expressions containing the squares of wavefunctions in the momentum space (**p**-space).

In Section 4, we describe calculations of the contribution of bound states with the use of the exact Fock equations for the Coulomb wavefunctions of the discrete spectrum in the **p**-space. The results are compared with those obtained using the Planck–Larkin equation.

Section 5 contains calculations of the contribution to the second virial coefficient of scattering states that arise in interactions in repulsive and attractive fields. The results are compared with those reported in [13].

Section 6 contains the general equation of state of a weakly nonideal hydrogen plasma on the Sun (where we must, of course, take into account a large number of chemical elements). In the central region with the plasma parameters $T \sim 1$ keV and $\rho \sim 150$ g/cm^3, we must also take into account the pressure of radiation in the medium. The corresponding equations are given.

In Section 7, we present the results of numerical calculations of the speed of sound and adiabatic compressibility performed for a hydrogen plasma along the solar trajectory within the framework of the so-called *S*-model [1].

2. THERMODYNAMIC PERTURBATION THEORY

The thermodynamic functions of a weakly nonideal hydrogen plasma can be calculated using the Matsubara technique [23, 24] or the Kadanoff–Baym–Keldysh formalism of nonequilibrium Green functions [25–27]. We will use the second method in several instances because it admits generalization to the nonequilibrium case and allows state broadening effects on the thermodynamic functions of a plasma to be included; these effects are responsible for the observation of discrete spectra.

Following [25], we calculate corrections to pressure P_0 of an ideal gas comprising noninteracting protons and electrons [28] to determine plasma pressure P,

$$P = P_0 + P_{\rm H} + P_{\rm exch} + P_{\rm D-H} + \delta P, \tag{4}$$

where $P_{\rm H}$ is the Hartree correction, $P_{\rm exch}$ is the exchange correction for electron–electron interaction, and $P_{\rm D-H}$ is the correction of the Debye–Hückel approximation (these terms are discussed in the present section). For higher order corrections δP that include the contribution of ladder diagrams, see Section 3.

We assume that the protons are nondegenerate. Their ideal-gas concentration (called activity, ζ_p) is related to the chemical potential μ_p and temperature ($\beta = T^{-1}$) in a grand canonical ensemble as

$$\zeta_p = 2\lambda\!\!\!^{-}_p{}^{-3}\exp(\beta\mu_p), \tag{5}$$

where

$$\lambda\!\!\!^{-}_p = \sqrt{\frac{2\pi\hbar^2}{m_p T}}$$

is the thermal de Broglie wavelength. The pressure of the ideal gas of the protons is

$$P_{0p} = T\zeta_p. \tag{6}$$

The electrons can be degenerate (for instance, in the center of the Sun, $n_e\lambda\!\!\!^{-}_e{}^3 \approx 0.6$), and their activity ζ_e will therefore be written [28] in terms of the ideal-gas concentration n_e^0 (valid in the general case of degenerate particles):

$$\zeta_e = n_e^0(\mu_e) = \frac{2}{\lambda\!\!\!^{-}_e{}^3}\frac{2}{\sqrt{\pi}}\int_0^\infty \frac{x^{1/2}dx}{\exp(x-y)+1},$$
$$\lambda\!\!\!^{-}_e = \sqrt{\frac{2\pi\hbar^2}{m_e T}}, \quad y = \frac{\mu_e}{T}. \tag{7}$$

Accordingly, the pressure of an ideal gas of degenerate electrons is given by [28]

$$P_{0e} = \frac{4}{3\sqrt{\pi}}T\frac{2}{\lambda\!\!\!^{-}_e{}^3}\int_0^\infty \frac{x^{3/2}dx}{\exp(x-y)+1}. \tag{8}$$

In this approximation, the electroneutrality condition is written as

$$\zeta_p = \zeta_e. \tag{9}$$

The Hartree correction to the ideal-gas approximation

$$P_0 = P_{0p} + P_{0e}$$

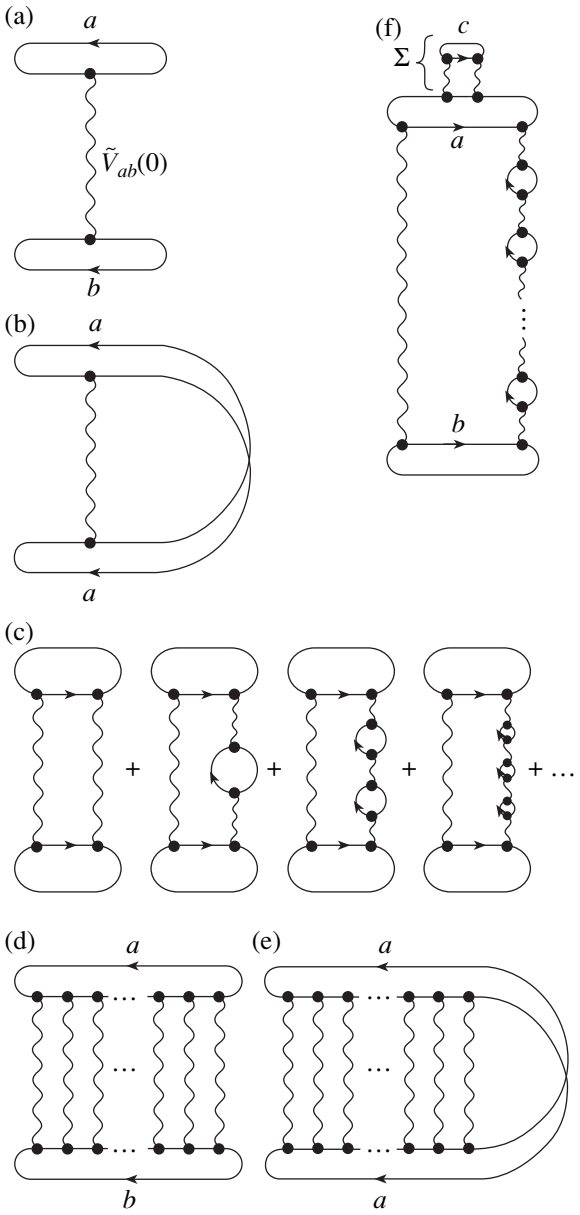


Fig. 1. Feynman diagrams in the Matsubara technique [23, 24] (solid lines are the Green function and wavy lines are the Fourier components of the interaction potential): (a) Hartree correction, see (10) and (109); (b) Hartree–Fock approximation (exchange interaction), Eq. (15); (c) Debye–Hückel approximation, Eqs. (16) and (43); (d) ladder diagram, Eqs. (58), (59), and (87); (e) ladder diagram (exchange interaction), Eqs. (79), (166), (167), and (170); and (f) diagram describing medium effects (taking into account energy corrections proper, Σ_c) on Debye–Hückel screening, Eq. (57).

(see Fig. 1a) has the form [7, 13, 28]

$$\frac{\delta\Omega_{\rm H}}{V} = \tilde{V}(0)(\zeta_e - \zeta_p)^2 \qquad (10)$$

for the Helmholtz thermodynamic potential $\Omega = -PV$ (V is the volume of the system). Here, $\tilde{V}(0)$ is the Fourier transform of the Coulomb interaction potential at zero momentum transfer. The regularization of the integral

$$\tilde{V}(\mathbf{q}) = \lim_{\kappa\to 0}\int\frac{e^2}{r}\exp(i\mathbf{q}\cdot\mathbf{r}-\kappa r)d\mathbf{r} = \frac{4\pi e^2}{q^2+\kappa^2}, \quad \mathbf{r} = (r_1, r_2, r_3), \quad |\mathbf{r}| = r, \quad d\mathbf{r} = dr_1dr_2dr_3, \qquad (11)$$

with the use of $\exp(-\kappa r)$, where $\kappa \longrightarrow 0$ is the regularization parameter, yields

$$\tilde{V}(0) = \frac{4\pi e^2}{\kappa^2}. \qquad (12)$$

This regularization of the Fourier component of the potential reduces (10) to zero because of electroneutrality condition (9). In the general case of a multicomponent plasma, condition (9) becomes

$$\zeta_e = \sum_k z_k\zeta_k \qquad (13)$$

(z_k is the charge of a particle of kind k in e units), and the generalization of (10) is obvious,

$$\frac{\delta\Omega_{\rm H}}{V} = \tilde{V}(0)\left(\zeta_e - \sum_k z_k\zeta_k\right)^2. \qquad (14)$$

In the next, Hartree–Fock, approximation (see Fig. 1b), we obtain the well-known [25, 28, 29] finite solution for the exchange correction caused by electron–electron interaction. For instance, for nondegenerate electrons in the first Born approximation, we have [28]

$$\frac{\delta\Omega_{\rm exch}}{V} = -\frac{2\pi e^2\hbar^2}{m_eT}\zeta_e^2. \qquad (15)$$

Here, ζ_e should be calculated using the nondegenerate limit [$\exp(-y) \gg 1$] in (7),

$$\zeta_e = 2\lambda_e^{-3}\exp(\mu_e/T).$$

We give a more general equation containing the sum over the entire Born series in Section 5.

The ring diagrams [30, 31] that follow next along the interaction potential (see Fig. 1c) yield the Debye–Hückel contribution (e.g., see [7–14, 24])

$$\frac{\delta\Omega_{\rm D-H}}{V} = -T\frac{\kappa_{\rm D}^3}{12\pi}. \qquad (16)$$

Here, κ_D is the inverse Debye radius [24, 32],

$$\kappa_D^2 = 4\pi e^2 \sum_k z_k^2 \left(\frac{\partial n_k}{\partial \mu_k}\right)_T. \quad (17)$$

Equation (17) was rigorously derived by Fradkin [32], who showed that the right-hand side of (17) contains derivatives of physical concentrations n_k with respect to chemical potentials. In perturbation theory, the activities can be approximated as $n_k^0(\mu_k) = \zeta_k$.

Recall that physical concentrations are related to chemical potentials as [28]

$$n_k = -\left(\frac{\partial(\Omega/V)}{\partial \mu_k}\right)_T. \quad (18)$$

The physical concentrations obey the traditional electroneutrality relation

$$n_e = \sum_k z_k n_k. \quad (19)$$

Attempts at relating the physical concentrations to activities according to (18) can give the following result: if plasma asymmetry, for instance, caused by electron exchange, is taken into account, the obtained ζ_k values do not satisfy the electroneutrality condition for activities [Eq. (9) or (13)], which leads to unbalanced divergence of Hartree contribution (10) to the thermodynamic potential. This divergence was removed in [7] in assuming that $\tilde{V}(0) = 0$.

Let us consider the simplest model of a nondegenerate plasma in which the exchange contribution $\delta\Omega_{\text{exch}}$ is ignored compared with the Debye–Hückel correction. This presupposes smallness of the parameter

$$\frac{(\lambda_e \kappa_D)^2}{\Gamma} \ll 1.$$

In this model,

$$\Omega = \Omega_0 + \delta\Omega_{D-H}. \quad (20)$$

Using activities instead of concentrations n_k in (17), we obtain from (18)

$$n_k = \zeta_k\left(1 + \frac{\Gamma_D}{2} z_k^2\right), \quad (21)$$

$$n_e = \zeta_e\left(1 + \frac{\Gamma_D}{2}\right). \quad (22)$$

If $z_k \neq 1$, Eqs. (21) and (22) with

$$n_e = \sum_k z_k n_k$$

yield

$$\zeta_e = \frac{n_e}{1 + \frac{\Gamma_D}{2}} = \frac{\sum_k z_k n_k}{1 + \frac{\Gamma_D}{2}} \neq \sum_k z_k \zeta_k = \sum_k \frac{z_k n_k}{1 + \frac{\Gamma_D}{2} z_k^2}. \quad (23)$$

That is, even within the framework of this simplest model, condition (13) is violated. For a singly charged hydrogen plasma, $\zeta_e = \zeta_p$ in this model, but this equality does not hold if exchange is taken into account. The technique that we use to correlate conditions (13) and (19), which is necessary for removing the divergence of Hartree term (14) as $\kappa \longrightarrow 0$, is as follows. Because we assume that conditions (13) and (19) should be satisfied simultaneously to prevent the divergence of the Hartree term, the concentrations should be determined according to (18) with differentiating the thermodynamic potential with respect to chemical potentials taking into account constraints imposed by (13). The n_e and ζ_e values are found from (19) and (13), respectively,

$$n_e + \sum_k n_k = \sum_k (z_k + 1) n_k = -\beta \sum_k \left(\frac{\partial(\Omega/V)}{\partial \zeta_k}\right)_T \zeta_k. \quad (24)$$

The differentiation and summation in (24) are only over the activities of the ions, and ζ_e in Ω is replaced by (13). For the model described by (20), Eq. (24) yields

$$\sum_k (z_k + 1) n_k = \sum_k (z_k + 1)\zeta_k + \frac{\Gamma_D}{2} \sum_k z_k (z_k + 1)\zeta_k. \quad (25)$$

It follows from (25) that [cf. (21)]

$$\zeta_k = \frac{n_k}{1 + \frac{\Gamma_D}{2} z_k}. \quad (26)$$

Using electroneutrality conditions (19) and (13) for physical concentrations and activities, respectively, and

taking into account (26), we easily obtain

$$n_e = \sum_k z_k n_k = \sum_k \zeta_k z_k \left(1 + \frac{\Gamma_D}{2} z_k\right) = \zeta_e + \frac{\Gamma_D}{2} \sum_k z_k^2 \zeta_k. \quad (27)$$

It follows from (27) that

$$\zeta_e = \sum_k z_k n_k - \frac{\Gamma_D}{2} \sum_k \frac{z_k^2 n_k}{1 + \frac{\Gamma_D}{2} z_k} = \sum_k \frac{z_k n_k}{1 + \frac{\Gamma_D}{2} z_k}. \quad (28)$$

That is, as distinct from (23), where independent definition (18) is used for the ions and electrons, definition (24) leads to correlated fulfillment of electroneutrality conditions (13) and (19) and removes the divergence of the Hartree term. To determine ζ_k explicitly through the physical concentrations n_k, we use the definition of the inverse Debye radius (17),

$$\kappa_D^2 = \frac{4\pi e^2}{T} \sum_k \frac{z_k(z_k + 1) n_k}{1 + \frac{\Gamma_D}{2} z_k}. \quad (29)$$

Here, the summation is over the kinds of ions. Multiplying (29) by $(e^2/T)^2$ yields the equation for

$$\Gamma_D(\zeta_k) \equiv \frac{\kappa_D e^2}{T}$$

in terms of the physical concentrations n_k,

$$\Gamma_D^2 = 4\pi \left(\frac{e^2}{T}\right)^3 \sum_k \frac{z_k(z_k + 1) n_k}{1 + \frac{\Gamma_D}{2} z_k}. \quad (30)$$

After determining $\Gamma_D(n_k)$ by (30), we can find the activities ζ_k and ζ_e by (26) and (28).

In the standard Debye–Hückel theory for the grand canonical ensemble [33, 34] with n_k and n_e given by (18) and (19), (30) is usually replaced by

$$\Gamma_D^2 = 4\pi \left(\frac{e^2}{T}\right)^3 \sum_m \frac{z_m^2 n_m}{1 + \frac{\Gamma_D}{2} z_m^2}. \quad (31)$$

Here, the summation is over the ions and electrons.

Solutions to (30) and (31) can be compared in the limit

$$\langle z \rangle \gg 1, \quad \frac{\Gamma_D}{2} (\langle z \rangle \gtrsim 1, \quad \Gamma_D \ll 1.)$$

(Strictly, the condition of the applicability of the Debye–Hückel approximation also requires that the inequality $(\Gamma_D/2) z_m \ll 1$ be satisfied.) It then follows from (30) that

$$\Gamma_D^N = 2 \frac{e^2}{T} (\pi n_e)^{1/3}. \quad (32)$$

Index "*N*" means that the "new" equation [Eq. (30)] is used. On the other hand, (31) yields

$$\Gamma_D^O = 2 \left(\frac{e^2}{T}\right)^{3/2} (\pi n_e)^{1/2} \quad (33)$$

(index "*O*" stands for old). The extrapolation to $\Gamma_D \gg 1$, which is, of course, unjustified, yields

$$\Gamma_D^O = \Gamma_D^N = 2 \frac{e^2}{T} (\pi n_e)^{1/3}, \quad (34)$$

according to both (30) and (31). If $\Gamma_D z_m^2 \ll 1$, (30) and (31) yield

$$\Gamma_D^N = \Gamma_D^O = \kappa \frac{e^2}{T}, \quad \kappa^2 = 4\pi \frac{e^2}{T} \sum_k z_k(z_k + 1) n_k.$$

Note that relations (21), (22), and (26–28) between concentrations and activities give the same equation of state for model (20),

$$P = T \left(\sum_k \zeta_k + \zeta_e\right) + T \frac{\Gamma_D}{3} \left(\sum_k z_k^2 \zeta_k + \zeta_e\right). \quad (35)$$

If the condition of electroneutrality for activities [Eq. (13)] is ignored, Eqs. (21) and (22) yield

$$P = T \sum_k (z_k + 1) n_k - \frac{\Gamma_D}{6} \sum_k z_k(z_k + 1) n_k \quad (36)$$

in the limit $\Gamma_D \ll 1$. Here, the summation is only over the ions and electroneutrality condition (19) is used. However, using condition (13) and solutions (26) and (28) in (35) yields the equation of state exactly in form (36). Relation (13) for solutions (26) and (28) means that

$$n_e \Delta\mu_e^N = \sum_k z_k n_k \Delta\mu_k. \quad (37)$$

The symbol "Δ" denotes the correction to the ideal-gas approximation μ_a^0,

$$\mu_a = \mu_a^0 + \Delta\mu_a$$

[see [28]; for instance, $\mu_e^0 = T\ln(n_e\chi_e^3/2)$]. It follows from (26) that

$$\Delta\mu_k^N = -T\frac{\Gamma_D}{2}z_k. \tag{38}$$

Using (38) in (37), we obtain

$$n_e\Delta\mu_e^N = -T\frac{\Gamma_D}{2}\sum_k z_k^2 n_k. \tag{39}$$

In the standard Debye–Hückel theory for the canonical ensemble (see [28]), we have

$$\Delta\mu_k^O = -T\frac{\Gamma_D}{2}z_k^2 \tag{40}$$

[compare (40) with solution (21)]. By analogy with (22), we find that

$$\Delta\mu_e^O = -T\frac{\Gamma_D}{2} \tag{41}$$

for the electrons. Relation (37) does not hold for solutions (40) and (41).

If the relation between concentrations [Eq. (19)] is taken into account, the relation between the chemical potentials and free energy F takes the form [cf. (24)]

$$\sum_k \Delta\mu_k n_k + n_e\Delta\mu_e = \sum_k n_k\frac{\partial(\Delta F/V)}{\partial n_k} + n_e\frac{\partial(\Delta F/V)}{\partial n_e}. \tag{42}$$

This equation is satisfied in the Debye–Hückel theory for both solutions (38), (39) and (40), (41).

In the region of the applicability of the theory ($z_k^2\Gamma_D \ll 1$), the equation of state (36) is insensitive to the selection of a solution to (24), but the selection in favor of condition (13) follows from the requirement of the absence of the divergence of (14) and other (higher order) terms of thermodynamic perturbation theory, such as the contribution of ladder diagrams in the expansion of Ω in powers of activity, whose finiteness is, in particular, provided by the vanishing of (14) (see Sections 3–5).

The Debye–Hückel approximation represented by the contribution of ring diagrams in the static limit [24] (see Fig. 1c) contains several corrections that can be obtained using the technique described in [25, 26] (particle spins are set equal to 1/2 for simplicity),

$$\frac{\Delta\Omega}{V} = -4\sum_{i,j}\int_0^1\frac{d\lambda}{2\lambda}\lambda^2\int\frac{d\mathbf{P}d\mathbf{q}d\mathbf{k}}{(2\pi)^9}\frac{\exp(-\beta(\varepsilon_k^{ij}-\varepsilon_q^{ij}))-1}{\varepsilon_q^{ij}-\varepsilon_k^{ij}} \times n_i\left(\frac{m_i}{M}\mathbf{P}+\mathbf{q}\right)\left(1-n_i\left(\frac{m_i}{M}\mathbf{P}+\mathbf{k}\right)\right) \times n_j\left(\frac{m_j}{M}\mathbf{P}-\mathbf{q}\right)\left(1-n_j\left(\frac{m_j}{M}\mathbf{P}-\mathbf{k}\right)\right) \times\frac{16\pi^2e^4z_i^2z_j^2}{(\mathbf{q}-\mathbf{k})^2[(\mathbf{q}-\mathbf{k})^2+4\pi e^2\lambda\Pi^R(\mathbf{q}-\mathbf{k},\mathbf{V}\cdot(\mathbf{q}-\mathbf{k}))]}, \tag{43}$$

where the integration in the λ parameter corresponds to the integration in charge $e^2 \mapsto e^2\lambda$; $\mathbf{P}$ is the total momentum of particles i and j; $\mathbf{q}$ and $\mathbf{k}$ are the wavevectors of the relative motion of the center of mass before and after interaction, respectively; $n_i(\mathbf{q})$ stands for the occupation numbers of the particles of kind i; m_i and z_i are the mass and charge (in elementary units) of a particle of kind i; $M = m_i + m_j$ is the total mass;

$$\varepsilon_k^{ij} = \frac{\hbar^2k^2}{2\mu}, \quad \mu = \frac{m_im_j}{m_i+m_j}$$

is the reduced mass of particles i and j; and

$$\mathbf{V} = \mathbf{P}/M, \quad \hbar|\mathbf{V}|q \sim \hbar|\mathbf{V}|\kappa_D < T.$$

This equation is for the general case that admits the inclusion of degeneracy and nonstatic corrections. Taking into account the contribution of two crosspieces to the sum of ring diagrams, which should be subtracted from the ladder diagrams (see Figs. 1d, 1e) that describe the interaction of a pair of particles in the continuum, corresponds to the replacement of $4\pi e^2\lambda\Pi^R$ in (43) in the denominator of the last multiplier by the regularization parameter κ^2 [cf. (11), (12)]. Here, Π^R is the retarded polarization operator [24–27] determined by the sum of the contributions of all the particles that participate in the screening of Coulomb interaction.

The static approximation implies the validity of the condition

$$\max\{T, \varepsilon_{Fi}\} \gg \hbar\omega_{pi}.$$

Here, ε_{Fi} is the Fermi energy of the particles of kind i and ω_{pi} is their plasma frequency [24, 28].

For instance, in the nondegenerate case ($n_i \ll 1$), it follows from (43) that

$$\frac{\Delta\Omega}{V} = -4\sum_{i,j}\frac{\exp(\beta(\mu_i+\mu_j))}{\varkappa_\Sigma^3}\int_0^1 d\lambda\frac{\lambda}{2} \times\int\frac{d\mathbf{q}d\mathbf{k}}{(2\pi)^6}\frac{\exp(-\beta\varepsilon_k^{ij})-\exp(-\beta\varepsilon_q^{ij})}{(\varepsilon_q^{ij}-\varepsilon_k^{ij})(\mathbf{q}-\mathbf{k})^2} \times\frac{16\pi^2e^4z_iz_j}{(\mathbf{q}-\mathbf{k})^2+4\pi e^2\lambda\Pi^R(\mathbf{q}-\mathbf{k},0)}. \quad (44)$$

Here,

$$\varkappa_\Sigma = \sqrt{\frac{2\pi\hbar^2}{MT}}.$$

The replacement of the multiplier

$$\frac{\exp(-\beta\varepsilon_k^{ij})-\exp(-\beta\varepsilon_q^{ij})}{\varepsilon_q^{ij}-\varepsilon_k^{ij}}$$

with

$$\beta\exp(-\beta\varepsilon_k^{ij})$$

(with respect to the $T \gg \hbar\kappa v_T^{ij}$ parameter, where $v_T^{ij} \sim \sqrt{2T/\mu}$ is the thermal velocity of relative motion) yields the approximation

$$\frac{\Delta\Omega}{V} = -\frac{T\kappa_D^4}{4\pi^2}\int_0^1\lambda d\lambda\int_0^\infty\frac{dq}{q^2+4\pi e^2\lambda\Pi^R(q)}. \quad (45)$$

Here,

$$\kappa_D^2 = 4\pi e^2\Pi^R(0)$$

is the square of the inverse Debye radius,

$$\Pi^R(q)\equiv\Pi^R(|\mathbf{q}|)\equiv\Pi^R(\mathbf{q},0).$$

In the nondegenerate case,

$$\kappa_D^2 = 4\pi e^2\beta\sum_i z_i^2 n_i.$$

Usually, the approximation is used in which $\Pi^R(q)$ is replaced by $\Pi^R(0)$. Equations (45) and (17) then yield (16).

We must clarify this point. In (45), $\Pi^R(q)$ should be expanded up to the second derivative with respect to q [35],

$$q^2+4\pi e^2\lambda\Pi^R(q) \longrightarrow q^2\left(1+2\pi e^2\lambda\frac{\partial^2\Pi^R}{\partial q^2}(0)\right)+4\pi e^2\lambda\Pi^R(0). \quad (46)$$

Let us introduce the value

$$a = \left|2\pi e^2\Pi''(0)\right|, \quad \Pi'' = \frac{\partial^2\Pi^R}{\partial q^2}.$$

For this value, we obtain [20, 35]

$$2\pi e^2\Pi''(0) = -\sum_i\frac{\pi\hbar^2e^2z_i^2}{m_i}\frac{\partial^2 n_i^0}{\partial\mu_i^2} + \frac{\pi\hbar^2}{9}\sum_i\frac{e^2z_i^2}{m_i}\frac{\partial^3(\overline{\varepsilon n}_i)}{\partial\mu_i^3}. \quad (47)$$

Here, n_i^0 is the ideal gas concentration and $\overline{\varepsilon n}_i$ is the averaging of the kinetic energy over occupation numbers for the particles of kind i,

$$\overline{\varepsilon n}_i = \int\frac{d\mathbf{p}}{(2\pi)^3}\frac{p^2}{2m_i}n_i(\mathbf{p}). \quad (48)$$

The Debye–Hückel correction then takes the form [20]

$$\frac{\Delta\Omega_{D-H}}{V} = -\frac{T\kappa_D^3}{12\pi}f(a). \quad (49)$$

Here,

$$f(a) = \frac{3}{2a^{3/2}}\{\arcsin\sqrt{a}-\sqrt{a(1-a)}\}. \quad (50)$$

Estimates for a hydrogen plasma give

$$2\pi e^2\Pi''(0) = -173\rho T^{-2}$$

for the nondegenerate case (ρ is the density of matter in g/cm^3 and T is the temperature in eV). In the center of the Sun, $f(a)$ differs from 1 in the third decimal place, which, together with the smallness of the Debye–Hückel correction ($\sim 10^{-2}$) compared with the ideal gas approximation, allows us to use approximation (16) in helioseismology problems [$f(a) = 1$].

For the "true" radius of screening of charge fluctuations $\Gamma_{scr} = \kappa^{-1}$, (46) yields

$$\kappa^2 = \frac{\kappa_D^2}{1 + 2\pi e^2 \Pi''(0)}. \quad (51)$$

It follows from the estimates of $\Pi''(0)$ given above that, at the plasma temperature $T \sim 1$ eV, the condition

$$1 + 2\pi e^2 \Pi''(0) = 0$$

is fulfilled at $n_e \approx 3 \times 10^{21}$ cm^{-3}. However, under these conditions, the nonideality parameter

$$\gamma = n^{1/3} e^2 / T \approx 2$$

and Eq. (47) for $\Pi''(0)$ must be refined. A charge density wave is formed and Debye screening disappears in the region where $\kappa^2 < 0$. The parameter with respect to which we must take into account corrections on the order of a corresponds to the value $(\chi_e \kappa_D)^2$, which is not too small in the center of the Sun [the closeness of $f(a)$ to one is determined by the smallness of the coefficient of this parameter]. The ratio between the thermal de Broglie wavelength of the electron and the Debye radius characterizes quantum effects in Debye screening, and the corresponding corrections to (16) and (49) [at $f(a) = 1$] are known as diffraction corrections [12]. They can easily be found from (44) in the approximation

$$4\pi e^2 \lambda \Pi^R(\mathbf{q} - \mathbf{k}, 0) = \lambda \kappa_D^2$$

taking into account exponential multipliers and energy denominators present in (44). In the nondegenerate case in the first approximation with respect to the $\chi\kappa$ parameter, (44) yields (see [12])

$$\frac{\Delta\Omega_{diff}}{V} = \frac{\pi^{3/2}}{4} T \left(\frac{e^2}{T}\right)^2 \times \left\{ \lambda_{ee} \zeta_e^2 + 2\zeta_e \sum_k \zeta_k z_k^2 \lambda_{ek} + \sum_{kj}{}' \zeta_k \zeta_j z_k^2 z_j^2 \lambda_{kj} \right\}. \quad (52)$$

Here,

$$\lambda_{ee} = \frac{\hbar}{\sqrt{m_e T}}, \quad \lambda_{ek} = \frac{\hbar}{\sqrt{2\mu_{ek} T}}, \quad \lambda_{kj} = \frac{\hbar}{\sqrt{2\mu_{kj} T}}.$$

Note also one more class of corrections to the Debye–Hückel approximation for the first time obtained in [8] (also see [9, 12]). Physically, these corrections arise because particles interacting with the medium ("dressed" particles, see Fig. 1f) participate in screening. Formally, this corresponds to the inclusion of energetic corrections proper to the Green functions present in polarization loops. The corrections calculated in [32] and [25, 27] are on the order of χ^2. Classical corrections arise when we take into account the energy diagram proper Σ_i^R [25, 26],

$$\mathrm{Re}\Sigma_i^R = \frac{1}{2}(\Sigma_i^{--} - \Sigma_i^{++}).$$

For particles of kind i,

$$\Sigma_i^{--}(\mathfrak{p}) = \int V_q \tilde{V}_q \Pi^{--}(\mathfrak{q}) G_i^{--}(\mathfrak{p} - \mathfrak{q}) \frac{d\mathfrak{q}}{(2\pi)^4}, \quad (53)$$

where

$$V_q = \frac{4\pi e^2}{q^2}, \quad \tilde{V}_q = \frac{4\pi e^2}{q^2 + \lambda\kappa_D^2},$$

and $\mathfrak{q} = (\mathbf{q}, q_4)$ is the four vector.

Calculations of operators of the type Π^{--} taking into account relations of the type

$$\Pi^{--} = \Pi^R - \Pi^{-+}$$

and the summation in calculating $\Pi^R(0)$ over all particle kinds give

$$\mathrm{Re}\Sigma_i^R = -\frac{1}{2}\lambda e^2 z_i^2 \kappa(\lambda) \quad (54)$$

in the nondegenerate case (cf. [14]). Here,

$$\kappa(\lambda) = \sqrt{\lambda}\kappa_D, \quad \kappa_D^2 = 4\pi \sum_j e^2 z_j^2 \left(\frac{\partial n_j}{\partial \mu_j}\right).$$

Using the obtained classical equation for the mass operator, we find the correction to the kinetic Green function in the form [27]

$$G_i^{-+} \equiv G_i^<(\mathfrak{p}) = -n_i(\omega)(G_i^R - (G_i^R)^*), \quad (55)$$

where

$$G_i^R = \frac{1}{\omega - \varepsilon_p - \Sigma_i^R(\mathfrak{p})}. \quad (56)$$

In (55),

$$n_i(\omega) = \{\exp(\beta(\omega - \mu_i)) + 1\}^{-1}$$

are the occupation numbers of the fermions. Using (54) in correction (43) and Eqs. (55) and (56), we obtain the

classical correction to the Debye–Hückel term (the index "*cl*" means that this result is classical in nature)

$$\frac{\delta\Omega_{cl}}{V} = -\frac{\pi}{3}T\left(\frac{e^2}{T}\right)^3\left(\sum_i \zeta_i z_i^4\right)\left(\sum_j \zeta_j z_j^2\right). \quad (57)$$

In (57), the summation is over all particle kinds. The coefficient $\pi/2$ was used in [8] instead of $\pi/3$. The same value is also found in [9]. In [13], this term is written with either $\pi/3$ or $\pi/2$ in different places. Our independent examination shows that the correct coefficient value is $\pi/3$. Note that this correction is small compared with the "large" logarithmic term discussed in Section 5.

3. THE LADDER APPROXIMATION IN SECOND VIRIAL COEFFICIENT CALCULATIONS

Let us consider the contribution of the δP value present in (4), see Fig. 1d. In the Matsubara technique [23, 24], $\delta\Omega/V = -\delta P$ is given by (the integration in λ implies the $e^2 \mapsto e^2\lambda$ replacement)

$$\frac{\delta\Omega_L}{V} = \frac{2}{\beta}\sum_0^1 \frac{d\lambda}{2\lambda}\int\frac{d\mathbf{p}}{(2\pi)^3}G_i(\mathbf{p}, \omega)\Sigma_i(\mathbf{p}, \omega). \quad (58)$$

Here, the summation is over the frequencies ω (or p_4); for fermions, $\omega = \pi T(2n + 1)$; the index "*L*" stands for ladder; $G_i(\mathbf{p}, \omega)$ is the Green function of a particle of kind i with momentum $\mathbf{p}$ and frequency ω in the Matsubara technique [24]; and the energy part proper $\Sigma_i(\mathbf{p}, \omega)$ can be written via the two-particle vertex Γ_{ij} found in the ladder approximation [29],

$$\Sigma_i(\mathfrak{p}) = \frac{2}{\beta}\sum_{j,k_4}\int\frac{d\mathbf{k}}{(2\pi)^3}G_j(\mathfrak{k}) \times \Gamma_{ij}\left(\frac{m_j\mathfrak{p} - m_i\mathfrak{k}}{m_i + m_j}, \frac{m_j\mathfrak{p} - m_i\mathfrak{k}}{m_i + m_j}; \mathfrak{p} + \mathfrak{k}\right). \quad (59)$$

For instance, for electron–proton interaction, $m_i = m_e$, $m_j = m_p$, $\mathfrak{p} = (\mathbf{p}, p_4) \equiv (\mathbf{p}, \omega)$ is the 4-momentum of the electron, and $\mathfrak{k} = (\mathbf{k}, k_4)$ is the 4-momentum of the proton. The value $\Gamma_{ij}(\mathfrak{q}, \mathfrak{q}'; \mathfrak{P})$ [$\mathfrak{q}$ and $\mathfrak{q}'$ are the relative 4-momenta before and after the reaction, respectively, and $\mathfrak{P} = \mathfrak{p} + \mathfrak{k} = (\mathbf{P}, P_4)$ is the total 4-momentum] can be written in the ladder approximation for small occupation numbers ($n \ll 1$) in the form [7, 29]

$$\Gamma_{ij}(\mathfrak{q}, \mathfrak{q}'; \mathfrak{P}) = (2\pi)^3\sum_n\left(iP_4 - \frac{\hbar^2P^2}{2M} - \frac{\hbar^2q^2}{2\mu} + \mu_i + \mu_j\right) \times \frac{\tilde\Psi_n(\mathbf{q})\tilde\Psi_n^*(\mathbf{q}')\left(E_n - \frac{\hbar^2q'^2}{2\mu}\right)}{iP_4 - \frac{\hbar^2P^2}{2M} - E_n + \mu_i + \mu_j}. \quad (60)$$

Here, $M = m_i + m_j$; $\mu = m_im_j/M$; μ_i and μ_j are the chemical potentials; E_n is the binding energy of the state with the principal quantum number n; the summation over n is generally over both discrete states with quantum numbers $\{n\} = (n, l, m)$, which are characterized by the wavefunctions $\tilde\Psi_n(\mathbf{q})$ of the relative motion of particles i and j (bound electron states in the field of protons), and continuous spectrum states; a tilde over a wavefunction denotes its Fourier component in the momentum space; and $P_4 = 2\pi nT$ is the fourth component of the total momentum. For the scattering states that correspond to the continuum, the sum over discrete states $\{n\}$ is replaced by the integration in momenta $\mathbf{k}$ that characterize the wavefunction at infinity (this is a plane wave in a short-range action field). Using the Schrödinger equation in the momentum representation and the completeness theorem for the wavefunctions ($\tilde V(\mathbf{q})$ is the Coulomb interaction potential in the momentum representation),

$$\left(E_n - \frac{\hbar^2q^2}{2\mu}\right)\tilde\Psi_n(\mathbf{q}) = \int\tilde V(\mathbf{q} - \mathbf{q}')\tilde\Psi_n(\mathbf{q}')\frac{d\mathbf{q}'}{(2\pi)^3}, \quad (61)$$

we can transform the scattering amplitude to

$$\Gamma_{ij}(\mathfrak{q}, \mathfrak{q}'; \mathfrak{P}) = \tilde V_{ij}(\mathbf{q} - \mathbf{q}') + (2\pi)^3 \times\sum_n\frac{\tilde\Psi_n(\mathbf{q})\tilde\Psi_n^*(\mathbf{q}')\left(E_n - \frac{\hbar^2q^2}{2\mu}\right)\left(E_n - \frac{\hbar^2q'^2}{2\mu}\right)}{iP_4 - \frac{\hbar^2P^2}{2M} - E_n + \mu_i + \mu_j}. \quad (62)$$

Using (58) and (59), let us write $\delta\Omega$ as

$$\frac{\delta\Omega}{V} = \sum_{i,j}\frac{4}{(2\pi)^6\beta^2}\sum_{q_4, P_4}\int_0^1\frac{d\lambda}{2\lambda}\int d\mathbf{q}d\mathbf{P} \times G_i\left(\frac{m_i}{M}\mathfrak{P} + \mathfrak{q}\right)G_j\left(\frac{m_j}{M}\mathfrak{P} - \mathfrak{q}\right)\Gamma_{ij}(\mathfrak{q}, \mathfrak{q}'; \mathfrak{P}). \quad (63)$$

The summation over the frequencies q_4 and P_4 and the integration in $d\mathbf{P}$ in (63) and (60) yield

$$\frac{\delta\Omega_L}{V} = \sum_{i,j}\zeta_i\zeta_j\lambda_{ij}^3\int_0^1\frac{d\lambda}{2\lambda}\int\frac{d\mathbf{q}}{(2\pi)^3} \times\sum_n\exp(-\beta E_n)\left(E_n - \frac{\hbar^2q^2}{2\mu}\right)|\tilde\Psi_n(\mathbf{q})|^2 \quad (64)$$

in the nondegenerate case (cf. [20]; we also use (62) for Γ_{ij} in what follows). Here, ζ_i stands for activities

[cf. (5)] and $\chi_{ij} = \sqrt{2\pi\hbar^2/\mu T}$. For an interacting proton and electron, we have $E_n = -\mathrm{Ry}/n^2$ in the discrete case and $E_k = \hbar^2k^2/2\mu$ in the continuum. For two particles i and j with the short-range interaction potential $V = \lambda V_0(\mathbf{r})$, the Beth–Uhlenbeck equation [21, 28] can be obtained from (64).

Let us use the relation that follows from the Schrödinger equation,

$$\int (E_n - \varepsilon_q)|\tilde{\Psi}_n(\mathbf{q})|^2 \frac{d\mathbf{q}}{(2\pi)^3} = \int V(\mathbf{r})|\Psi_n(\mathbf{r})|^2 d\mathbf{r}. \tag{65}$$

By virtue of the well-known quantum-mechanical theorem [36] and taking into account that

$$\frac{V(\mathbf{r})}{\lambda} = \frac{\partial V}{\partial \lambda} = \frac{\partial \hat{\mathcal{H}}}{\partial \lambda}, \tag{66}$$

where $\hat{\mathcal{H}}$ is the Hamiltonian of the relative motion of particles i and j, we have

$$\frac{\partial E_n}{\partial \lambda} = \left\langle n \left| \frac{\partial \hat{\mathcal{H}}}{\partial \lambda} \right| n \right\rangle \tag{67}$$

for bound states. Taking into account (65)–(67) and using (64), we obtain the contribution of discrete spectrum states in the form

$$\frac{\delta\Omega_{ij}^b}{V} = -\zeta_i\zeta_j\chi_{ij}^3 T\sum_{\{n\}}(\exp(-\beta E_n) - 1) \tag{68}$$

(the index "b" denotes bound states). After the summation over l and m and the substitution $E_n = -\mathrm{Ry}/n^2$, (68) diverges as

$$\sum_n (1) \sim \zeta(0),$$

where $\zeta(s)$ is the Riemann ζ-function,

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s}. \tag{69}$$

Equation (64) can also be used to obtain an explicit equation for the contribution of scattering states of particles i and j in a short-range potential.

The Schrödinger equation for the complex-conjugate wavefunction $\Psi_k^*(\mathbf{r})$,

$$\left[-\frac{\hbar^2}{2\mu}\Delta + \lambda V_0(\mathbf{r}) - \varepsilon_k\right]\Psi_k^*(\mathbf{r}) = 0, \tag{70}$$

can be differentiated with respect to λ to obtain the equation conjugate to (70),

$$\left[-\frac{\hbar^2}{2\mu}\Delta + \lambda V_0 - \varepsilon_k\right]\frac{\partial \Psi_k}{\partial \lambda} = -V_0\Psi_k. \tag{71}$$

Let us multiply (70) by $\partial\Psi_k/\partial\lambda$ and (71) by Ψ_k^*, subtract the latter result from the former, and integrate the difference over volume. This yields

$$J_c = \int V_0(\mathbf{r})|\Psi_k(\mathbf{r})|^2 d\mathbf{r} = -\frac{\hbar^2}{2\mu}\int\left(\frac{\partial\Psi_k}{\partial\lambda}\Delta\Psi_k^* - \Psi_k^*\Delta\frac{\partial\Psi_k}{\partial\lambda}\right)d\mathbf{r}. \tag{72}$$

The right-hand side of (72) in a short-range potential can be transformed into an integral over the surface using the asymptotic behavior of the wavefunction ($r \longrightarrow \infty$)

$$\Psi_k(\mathbf{r}) = \exp(i\mathbf{k}\cdot\mathbf{r}) + f_k(\theta)\frac{\exp(ikr)}{r}. \tag{73}$$

Here,

$$\mathbf{k}\cdot\mathbf{r} = kr\cos\theta = kz,$$

and f_k is the scattering amplitude.

Representation (73) is inapplicable to the Coulomb potential because of plane wave distortions at arbitrarily large distances. If the scattering amplitude is written through the scattering phases δ_l [36], (72) and (73) allow us to obtain

$$J_c = -\frac{\hbar^2}{2\mu}\frac{4\pi}{k}\sum_{l=0}^{\infty}(2l+1)\frac{\partial\delta_l}{\partial\lambda}. \tag{74}$$

Combining (68) and (74) yields

$$\frac{\Delta\Omega_{ij}^L}{V} = -\zeta_i\zeta_j\chi_{ij}^3\left[T\sum_n(\exp(-\beta E_n) - 1) + \sum_{l=0}^{\infty}(2l+1)\int_0^{\infty}\frac{dE}{\pi}\delta_l(E)\exp(-\beta E)\right]. \tag{75}$$

The Levinson theorem that relates the finite number of bound states n_l to momentum l via scattering

phases [22],

$$n_l = \frac{1}{\pi}[\delta_l(0) - \delta_l(\infty)], \tag{76}$$

is valid in a short-range potential. We can use (76) to transform (75) to the Beth–Uhlenbeck equation [21, 28].

Unfortunately, Eqs. (68), (75), and (76) diverge in the Coulomb case, and the use of the Beth–Uhlenbeck equation as a point of departure [10–13] therefore rests on an insecure (mystic) foundation.

Note that (64) can be generalized to include exchange of identical particles [29] (see Fig. 1e). Only a continuous spectrum exists in the repulsive field of identical particles, and the square of the modulus of the wavefunction $|\tilde{\Psi}_n(\mathbf{q})|^2$ should be replaced as

$$|\tilde{\Psi}_k(\mathbf{q})|^2 \longrightarrow \tilde{\Psi}_k^*(\mathbf{q})\left[\Psi_k(\mathbf{q}) - \frac{1}{2}\Psi_k(-\mathbf{q})\right]. \tag{77}$$

The second term in (77) gives an additional exchange contribution, which can be represented in the **r**-space using the identity

$$\int (E_k - \varepsilon_q)\Psi_k^*(\mathbf{q})\Psi_k(-\mathbf{q})\frac{d\mathbf{q}}{(2\pi)^3} = \int \Psi_k^*(\mathbf{r})V(\mathbf{r})\Psi_k(-\mathbf{r})d\mathbf{r}. \tag{78}$$

Accordingly, repeating the derivation of (75), we can use (78) to obtain a generalization of the Beth–Uhlenbeck equation to the case of identical particles (e.g., electrons),

$$\frac{\delta\Omega_{ee}^L}{V} = -\frac{\zeta_e^2}{2}\lambda_{ee}^3\sum_{l=0}^{\infty}(2l+1) \times \int_0^{\infty}\frac{\delta_l(E)}{\pi}\left[1 - \frac{(-1)^l}{2}\right]\exp(-\beta E)dE. \tag{79}$$

We must, for scattering states, specify the meaning of the square of the modulus of a wavefunction in the **p**-representation. Indeed, the short-range potential $\Psi_k(\mathbf{q})$ contains the three-dimensional δ-function $\delta(\mathbf{k} - \mathbf{q})$ because of the presence of a plane wave in the boundary condition.

Let us consider the right-hand side of (65) (this value will be denoted by J) and use the integral representation of the wavefunction [36]

$$\Psi_k(\mathbf{r}) = \exp(i\mathbf{k}\cdot\mathbf{r}) + \frac{2\mu}{\hbar^2}\int G_0(\mathbf{r}, \mathbf{r}')V(\mathbf{r}')\Psi_k(\mathbf{r}')d\mathbf{r}'. \tag{80}$$

Here, $G_0(\mathbf{r}, \mathbf{r}')$ is the Green function that describes the free particle,

$$G_0(\mathbf{r}, \mathbf{r}') = -\frac{1}{4\pi}\frac{\exp(ik|\mathbf{r} - \mathbf{r}'|)}{|\mathbf{r} - \mathbf{r}'|} = \int\frac{d\mathbf{p}}{(2\pi)^3}\frac{\exp(i\mathbf{p}\cdot(\mathbf{r} - \mathbf{r}'))}{p^2 - k^2 - i0}. \tag{81}$$

Using the definition of the scattering amplitude outside the mass surface

$$f(\mathbf{p}, \mathbf{k}) = -\frac{\mu}{2\pi\hbar^2}\int\exp(-i\mathbf{p}\cdot\mathbf{r})V(\mathbf{r})\Psi_k(\mathbf{r})d\mathbf{r} \tag{82}$$

(here, $|\mathbf{p}| \neq |\mathbf{k}|$ in the general case) and Eqs. (80) and (81), we obtain the right-hand side of (65) in the form

$$J = -\frac{2\pi\hbar^2}{\mu}\left[f(\mathbf{k}, \mathbf{k}) + 4\pi\int\frac{d\mathbf{p}}{(2\pi)^3}\frac{|f(\mathbf{p}, \mathbf{k})|^2}{p^2 - k^2 + i0}\right]. \tag{83}$$

Applying the optical theorem [36], we can write the term in square brackets also in the form

$$\mathrm{Re}\, f(\mathbf{k}, \mathbf{k}) + 4\pi P\int\frac{d\mathbf{p}}{(2\pi)^3}\frac{|f(\mathbf{p}, \mathbf{k})|^2}{p^2 - k^2}. \tag{84}$$

The symbol P means that the integral is calculated in the sense of its principal value.

Equations (80)–(82) also give $\tilde{\Psi}_k(\mathbf{q})$ in the form

$$\tilde{\Psi}_k(\mathbf{q}) = \int\exp(-i\mathbf{q}\cdot\mathbf{r})\Psi_k(\mathbf{r})d\mathbf{r} = (2\pi)^3\delta(\mathbf{k} - \mathbf{q}) + \frac{4\pi f(\mathbf{q}, \mathbf{k})}{q^2 - k^2 - i0}. \tag{85}$$

Using (85) in (65) and taking into account that

$$\delta(\mathbf{k} - \mathbf{q})(q^2 - k^2) = 0,$$

let us consider the left-hand side of (65). We obtain

$$\int\frac{d\mathbf{q}}{(2\pi)^3}\tilde{\Psi}_k^*(\mathbf{q})[\tilde{\Psi}_k(\mathbf{q})(\varepsilon_k - \varepsilon_q)] = \int\frac{d\mathbf{q}}{(2\pi)^3}\left\{(2\pi)^3\delta(\mathbf{k} - \mathbf{q}) + \frac{4\pi f^*(\mathbf{q}, \mathbf{k})}{q^2 - k^2 + i0}\right\} \times\left(-\frac{2\pi\hbar^2}{\mu}\right)f(\mathbf{q}, \mathbf{k}) = -\frac{2\pi\hbar^2}{\mu}\left[f(\mathbf{k}, \mathbf{k}) + 4\pi\int\frac{d\mathbf{q}}{(2\pi)^3}\frac{|f(\mathbf{q}, \mathbf{k})|^2}{q^2 - k^2 + i0}\right]; \tag{86}$$

that is, the method for associative multiplication of wavefunctions in the **p**-space specified above brings (86) in coincidence with (84).

However, if we first multiply $\Psi_k^*(\mathbf{q})$ by $\Psi_k(\mathbf{q})$ taking into account (85), we encounter problems with the determination of the square of the three-dimensional δ-function. Even if $[\delta(\mathbf{k}-\mathbf{q})]^2(q^2-k^2)$ is assumed to be zero, which is generally not obvious (but inevitable, because otherwise, the left-hand side of the identity remains finite while the right-hand side vanishes as the charge tends to zero), there appears the double product of the δ-function and the second term in (85), which entails the appearance of the superfluous factor 2 in the $f(\mathbf{k}, \mathbf{k})$ term in (86) and the violation of identity (65).

Equations of type (80) become invalid in the Coulomb field and, as has been mentioned, the representation of the scattering amplitude in form (60), which leads to (64), is accompanied by divergences.

If we use representation (62) for $\Gamma_{ij}(\mathfrak{q}, \mathfrak{q}; \mathfrak{P})$ present in (63) to determine the contribution of ladder diagrams, it becomes clear that the contribution of $\tilde{V}_{ij}(0) \sim e^2/\kappa^2$ [see (11), (12)] vanishes because of neutrality relation (13) for activities. After the summation over q_4 and P_4 in (63) with the use of the contribution of the second term of (62), we obtain (see Fig. 1d)

$$\frac{\Delta\Omega_L}{V} = \sum_{i,j}\zeta_i\zeta_j\lambda_{ij}^3\int_0^1\frac{d\lambda}{2\lambda}\int\frac{d\mathbf{q}}{(2\pi)^3} \times\sum_n\left(E_n-\frac{\hbar^2q^2}{2\mu}\right)\left|\tilde{\Psi}_n(\mathbf{q})\right|^2(\exp(-\beta E_n)-\exp(-\beta\varepsilon_q)). \quad (87)$$

Here, for the *e*–*p* interaction, we must sum over the discrete spectrum (over bound states) and integrate over scattering states characterized by the index **k**. For *e*–*e* and *p*–*p* interactions, only integration makes sense.

We show in Sections 4 and 5 that (87) yields finite solutions for the contributions of both bound and scattering states after the subtraction of the contribution of two crosspieces [$\sim e^4\beta/\kappa$, where $\kappa \longrightarrow 0$ is the regularization parameter, cf. (11), (12)] already taken into account in the summation of ring diagrams. In a hydrogen plasma, the total contribution of the logarithmic term proportional to $(e^2/T)^3\ln(m_p/m_e)$, which does not require being cut off at the Debye radius, is also finite.

4. THE CONVERGING CONTRIBUTION OF BOUND STATES TO THE SECOND VIRIAL COEFFICIENT

The contribution of bound states to the second virial coefficient can be calculated using the exact result obtained by Fock [37] for the wavefunctions of the nonrelativistic hydrogen atom in the momentum representation,

$$\frac{1}{(2\pi)^3}\sum_{l,m}\left|\tilde{\Psi}_{n,l,m}(\mathbf{q})\right|^2 = \frac{8}{\pi^2a_0^5n^3(q^2+p_n^2)^4}. \quad (88)$$

Here, $p_n = (a_0 n)^{-1}$ and $a_0 = \hbar^2/\mu e^2\lambda$ is the Bohr radius corresponding to the current charge $e^2\lambda$. As

$$E_n \equiv -\frac{\hbar^2p_n^2}{2\mu},$$

the part of (87) that corresponds to bound states can, taking into account (88), be written as

$$\frac{\delta\Omega^{BS}}{V} = -4\zeta_e\zeta_p\frac{\mu^4\lambda_{ep}^3e^{10}}{\pi^2\hbar^8}\sum_{n=1}^{\infty}\frac{1}{n^3} \times\int_0^1 d\lambda\lambda^4\left[\exp\left(\frac{\lambda^2X}{n^2}\right)\int\frac{d\mathbf{q}}{(q^2+p_n^2)^3} -\int\frac{d\mathbf{q}}{(q^2+p_n^2)^3}\exp\left(-\beta\frac{\hbar^2q^2}{2\mu}\right)\right]. \quad (89)$$

Here, $X = \beta\mathrm{Ry}$. We find from (89) that

$$\int\frac{d\mathbf{q}}{(q^2+p_n^2)^3} = \frac{\pi^2}{4p_n^3}, \quad (90)$$

$$\int\frac{d\mathbf{q}}{(q^2+p_n^2)^3}\exp\left(-\beta\frac{\hbar^2q^2}{2\mu}\right) = c_n(X), \quad \beta\frac{\hbar^2q^2}{2\mu} = \lambda_{ep}^2q^2, \quad \lambda_{ep} = \frac{\hbar}{\sqrt{2\mu T}}. \quad (91)$$

After transformations, the explicit equation for $c_n(X)$ takes the form

$$c_n(X) = 4\pi\int_0^{\infty}\frac{q^2dq}{(q^2+p_n^2)^3}\exp(-q^2\lambda_{ep}^2) = -4\pi\frac{\partial}{\partial z}\int_0^{\infty}\frac{dq}{q^2+z}\exp(-q^2\lambda_{ep}^2) -4\pi\frac{z}{2}\frac{\partial^2}{\partial z^2}\int_0^{\infty}\frac{dq}{q^2+z}\exp(-q^2\lambda_{ep}^2). \quad (92)$$

After the differentiation with respect to *z*, we must set $z = p_n^2$ here.

Consider the integral [38]

$$J_0 = \int_0^\infty \frac{dq}{q^2+z}\exp(-q^2\lambda_{ep}^2) = [1-\Phi(\lambda_{ep}p_n)]\frac{\pi}{2p_n}\exp(\lambda_{ep}^2p_n^2) = \Gamma\left(\frac{1}{2},\lambda_{ep}^2z\right)\frac{1}{2}\sqrt{\frac{\pi}{z}}\exp(\lambda_{ep}^2z). \quad (93)$$

Here, the incomplete gamma function $\Gamma\left(\frac{1}{2},z\right)$ is defined as follows:

$$\Gamma\left(\frac{1}{2},z\right) = \int_z^\infty \frac{e^{-t}}{\sqrt{t}}dt, \quad (94)$$

$$1-\Phi(\lambda_{ep}p_n) = \frac{1}{\sqrt{\pi}}\Gamma\left(\frac{1}{2},\frac{\lambda^2X}{n^2}\right). \quad (95)$$

Using (93) and (94) and performing the differentiation in (92), we obtain

$$c_n(X) = \pi^{3/2}\left[\frac{1}{4}a_0^3n^3\exp\left(\frac{\lambda^2X}{n^2}\right)\Gamma\left(\frac{1}{2},\frac{\lambda^2X}{n^2}\right) - \frac{a_0\beta\hbar^2}{2\mu}n\exp\left(\frac{\lambda^2X}{n^2}\right)\Gamma\left(\frac{1}{2},\frac{\lambda^2X}{n^2}\right) + \frac{\hbar a_0^2n^2}{2}\sqrt{\frac{\beta}{2\mu}} + \left(\frac{\beta\hbar^2}{2\mu}\right)^{3/2} - \frac{\beta^2\hbar^4}{4\mu^2a_0n}\exp\left(\frac{\lambda^2X}{n^2}\right)\Gamma\left(\frac{1}{2},\frac{\lambda^2X}{n^2}\right)\right]. \quad (96)$$

In the limit of $X/n^2 \ll 1$, (96) transforms into

$$c_n(X) = \frac{\pi^2a_0^3n^3}{4} - \frac{3}{8}\frac{\pi^2a_0n\beta\hbar^2}{\mu} + \frac{8}{3}\pi^{3/2}\left(\frac{\beta\hbar^2}{2\mu}\right)^{3/2} + \ldots. \quad (97)$$

It is easy to see that the contribution of the first term in (89) given by (90) and the first term of expansion (97) cancel each other out at $n \gg 1$.

Using (89), (90), (91), and (96), we obtain

$$\frac{\Delta\Omega^{BS}}{V} = -T\zeta_e\zeta_p\lambda\!\!\!^{-3}_{ep}\sum_{n=1}^\infty n^2\exp\left(\frac{I}{n^2T}\right)F_n(\beta) \quad (98)$$

after the integration in λ. Here, $F_n(\beta)$ has the form

$$F_n(\beta) = 1-\exp(-\beta I_n)\times\left[4-\frac{6}{\sqrt{\pi}}(\beta I_n)^{1/2}+\frac{4}{\sqrt{\pi}}(\beta I_n)^{3/2}\right] + \frac{\Gamma\left(\frac{1}{2},\beta I_n\right)}{\sqrt{\pi}}[3-4\beta I_n+4(\beta I_n)^2], \quad I_n = I/n^2. \quad (99)$$

Asymptotically, at $n \gg 1$, (99) yields

$$F_n(\beta) \sim 2(\beta I_n)^2. \quad (100)$$

This is four times larger than according to the Planck–Larkin equation,

$$F_n^{P-L}(\beta) = 1-\exp(-\beta I_n) - \beta I_n\exp(-\beta I_n) \longrightarrow \frac{1}{2}(\beta I_n)^2. \quad (101)$$

Expansion (97) can be used to write (89) in the form

$$\frac{\delta\Omega^{BS}}{V} = -\zeta_e\zeta_pT\lambda\!\!\!^{-3}_{ep}X\sum_{n=1}^\infty 2\int_0^1\lambda d\lambda\left[\exp\left(\frac{\lambda^2X}{n^2}\right) - 1 - 3\frac{\lambda^2X}{n^2} - \frac{32}{3}\frac{\lambda^3}{\sqrt{\pi}n^3}X^{3/2}+\ldots\right]. \quad (102)$$

Here, one that corresponds to the first term of expansion (97) is subtracted from the exponential function. The integration of (102) in λ yields

$$\frac{\delta\Omega^{BS}}{V} = -\zeta_e\zeta_pT\lambda\!\!\!^{-3}_{ep}\sum_{n=1}^\infty n^2\left(\exp\left(\frac{X}{n^2}\right)-1-\frac{X}{n^2} + \frac{3}{2}\frac{X^2}{n^4}-\frac{64}{15}\frac{X^{5/2}}{\sqrt{\pi}n^5}+\ldots\right). \quad (103)$$

We see that the first three terms under the sum sign in (103) correspond to the Planck–Larkin equation. The remainder of the expansion of the exponential function [see (99)] $X^2/2n^4$ plus the fourth term of sum (103) gives asymptotics (100).

Let us write the contribution of the bound states in the form

$$\frac{\delta\Omega^{BS}}{V} = -\zeta_e\zeta_pT\lambda\!\!\!^{-3}_{ep}\Sigma^{BS}. \quad (104)$$

Using the expansion in X, we obtain for (99)

$$\Sigma_{SRM}^{BS} = \sum_{k=4}^{\infty} \zeta(k-2)\frac{(-\sqrt{X})^k}{\Gamma\left(\frac{k}{2}+1\right)}(k-2)^2 + \sum_{k=1}^{\infty} \zeta(2k+1)\frac{X^{k+3/2}}{\Gamma\left(k+\frac{5}{2}\right)}. \quad (105)$$

At the same time, it follows from the Planck–Larkin equation that

$$\Sigma_{\text{P–L}}^{BS} = \sum_{k=2}^{\infty} \zeta(2k-2)\frac{X^k}{\Gamma(k+1)} = \sum_{n=4,6,8,\ldots} \zeta(n-2)\frac{X^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)}. \quad (106)$$

Let us introduce the notation

$$\Sigma_{\text{P–L}}^{BS} - \Sigma_{SRM}^{BS} = \Sigma_c \quad (107)$$

for the difference between $\Sigma_{\text{P–L}}^{BS}$ (106) and Σ_{SRM}^{BS} (105). This difference is given by

$$\Sigma_c = \sum_{k=2}^{\infty} \zeta(2k-2)\frac{X^k}{\Gamma(k+1)} - \sum_{k=4}^{\infty} \zeta(k-2)\frac{(-1)^k X^{k/2}}{\Gamma\left(\frac{k}{2}+1\right)}(k-2)^2 - \sum_{k=1}^{\infty} \zeta(2k+1)\frac{X^{k+3/2}}{\Gamma\left(\frac{k}{2}+1\right)}. \quad (108)$$

We show in Section 5, where the contribution of scattering states is considered, that (105) together with (87) calculated properly for the continuous spectrum coincides with the Σ^{tot} value for the total e–p interaction given in [13].

Using the completeness theorem for the functions including the discrete and continuous spectra, the subtrahend in (87) can, for instance, for e–p interaction, be written as (see Fig. 1a)

$$-\zeta_e\zeta_p\lambda_{ep}^3\int_0^1\frac{d\lambda}{\lambda}\int\frac{d\mathbf{q}}{(2\pi)^3}\exp(-\beta\varepsilon_q)\left(\sum_n + \int\frac{d\mathbf{k}}{(2\pi)^3}\right) \times \int\tilde{V}(\mathbf{q}-\mathbf{q}')\tilde{\Psi}^*_{\left\{\begin{smallmatrix}n\\k\end{smallmatrix}\right\}}(\mathbf{q})\tilde{\Psi}_{\left\{\begin{smallmatrix}n\\k\end{smallmatrix}\right\}}(\mathbf{q}')\frac{d\mathbf{q}'}{(2\pi)^3} = -\zeta_e\zeta_p\tilde{V}_{ep}(0), \quad (109)$$

that is, this subtrahend in (87) exactly balances the first term in representation (62) for $\Gamma_{ep}(\mathfrak{q}, \mathfrak{q}'; \mathfrak{P})$, which, by virtue of neutrality condition (13) for activities, vanishes after the summation over all pair interactions. It follows that the finite partition function (98), (99) obtained above corresponds to taking into account the summation over bound states only in counterterm (109). The integral over scattering states in (87) also becomes finite, but only after the subtraction of the contribution of the ladder section with two crosspieces included in the summation of the contribution of ring diagrams.

Consider a generalization of (89)-, (102)-type equations for the contribution of degenerate electrons to $\delta\Omega^{BS}$ ($m_e \ll m_p$),

$$\frac{\delta\Omega^{BS}}{V} = -\zeta_p\frac{32}{\pi}TX\sum_{n=1}^{\infty}2\int_0^1\lambda d\lambda \times\int_0^{\infty}\frac{t^2dt}{(1+t^2)^3}\frac{\exp\left(\frac{\lambda^2X}{n^2}(1+t^2)\right)-1}{\exp\left(\frac{\lambda^2X}{n^2}t^2-y\right)+1}. \quad (110)$$

At $\lambda^2X/n^2 \ll 1$ and $y = \mu_e/T > 0$, (110) gives

$$\frac{\delta\Omega^{BS}}{V} = -\zeta_p\frac{2\pi^2}{3}T\left(\frac{\text{Ry}}{T}\right)^2. \quad (111)$$

Equation (111) was derived on the assumption that the condition

$$\frac{\exp y}{2y^{5/2}}X^{5/2} \lesssim 1 \quad (112)$$

was satisfied. At $X \gg 1$ and $X > y > 0$, which corresponds to the definition $I > \mu_e$ in the degenerate case, (110) gives

$$\frac{\delta\Omega^{BS}}{V} = -2\zeta_pT\exp y\exp\frac{I}{T}. \quad (113)$$

The condition $X > y$ and contribution (113) of bound states correspond to anomalously high concentrations of "atomic" quasi-particles, and, because of strong interaction between them, both are inapplicable. In addition, the scattering amplitude Γ was obtained ignoring effects of the Pauli-blocking type [29].

Note that equations of type (87) are obtained in the Keldysh technique (see [25–27]),

$$\delta P = \sum_a (2S_a + 1)\int_0^1 \frac{d\lambda}{2\lambda}\int \frac{d\mathbf{p}}{(2\pi)^3} \times \int \frac{d\omega d\omega'}{(2\pi)^2}\frac{\Sigma_a^>(\mathbf{p}, \omega)G_a^<(\mathbf{p}, \omega')}{\omega - \omega'}(1 - \exp(-\beta(\omega - \omega'))). \quad (114)$$

Symbolically expressing the $\Sigma^>$ value through the imaginary part of the scattering amplitude Γ_{ep} [25],

$$\Sigma_e^> \sim \int \mathrm{Im}\Gamma_{ep}(\omega + \Omega)\cdot G_i^<(\mathbf{p}', \Omega)\frac{d\mathbf{p}' d\Omega}{(2\pi)^4}, \quad (115)$$

we obtain Γ_{ep} in the form [cf. (62)]

$$\mathrm{Im}\Gamma_{ep} \sim \sum_n |\tilde{\Psi}_n(\mathbf{q})|^2 (E_n - \varepsilon_q)^2 \times \delta_\gamma\left(\omega + \Omega - E_n - \frac{p^2}{2M}\right), \quad (116)$$

where $\delta_\gamma(x)$ is the Lorentz contour, which becomes the δ-function in the limit of zero width $\gamma \longrightarrow 0$. Equations (114)–(116) give (87) and (110). At the same time, this approach allows us to take into account atomic state broadening effects with the same accuracy as can be attained in calculating spontaneous radiation of a system of charged particles with discrete–discrete transitions. If Stark microfields that cause level shifts and statistical line broadening are taken into account, this approach also contains the effects of microfield-induced state decay [5, 17, 20].

Note that the atomic partition function in the lowest order in interaction (up to the second virial coefficient) converges starting with principal quantum number values $n_{\max} \sim \sqrt{\beta \mathrm{Ry}}$. This means that the contribution of bound states to pressure is given by an equation of the type Σ_{SRM} (or $\Sigma_{\text{P–L}}$). In contrast, discrete states can make an observable contribution to other physical effects, for instance, spontaneous radiation of an equilibrium plasma (the corresponding partition function converges) at $n > n_{\max}$, up to states at which their Stark width is comparable to the energy of their ionization.

5. THE CONVERGING CONTRIBUTION OF SCATTERING STATES TO THE SECOND VIRIAL COEFFICIENT

The contribution of continuous spectrum states to equations of type (87) can be calculated by determining the Fourier components of the wavefunctions that describe scattering of charged particles by each other. The system of Coulomb wavefunctions represented by the summation over orbital momenta [36] can conveniently be used. For instance, for the attractive field,

$$\Psi_k(\mathbf{r}) = \frac{1}{(2\pi)^{3/2}}\exp(\pi\tilde{\xi}/2)\Gamma(1 - i\tilde{\xi}) \times \exp(i\mathbf{k}\cdot\mathbf{r})F(i\tilde{\xi}, 1, i(kr + \mathbf{k}\cdot\mathbf{r})). \quad (117)$$

Here,

$$\tilde{\xi} = (a_0 k)^{-1} = \frac{\mu e^2 \lambda}{\hbar^2 k},$$

and $F(\alpha, \beta, z)$ is the degenerate hypergeometric function. The Fourier transform of (117) can be found using the linear regularization procedure,

$$\tilde{\Psi}_k(\mathbf{q}) = \int d\mathbf{r}\exp(-i\mathbf{q}\cdot\mathbf{r} - \kappa_1 r)\Psi_k(\mathbf{r}). \quad (118)$$

The κ_1 parameter must turn to zero at the end of calculations. Equations (117) and (118) give [39]

$$\tilde{\Psi}_k(q) = \frac{1}{(2\pi)^{3/2}}\exp(\pi\tilde{\xi}/2)\Gamma(1 - i\tilde{\xi})J_{\kappa_1}. \quad (119)$$

Here, J_{κ_1} has the form

$$J_{\kappa_1} = \left\{\frac{2\pi(1 - i\tilde{\xi})\kappa_1}{\left[\frac{(\mathbf{q} - \mathbf{k})^2}{2} + \frac{\kappa_1^2}{2}\right]^2} + \frac{2\pi\tilde{\xi}(k + i\kappa_1)}{\left[\frac{(\mathbf{q} - \mathbf{k})^2}{2} + \frac{\kappa_1^2}{2}\right]\left[\frac{q^2 - k^2 + \kappa_1^2}{2} - ik\kappa_1\right]}\right\} \times \exp\left\{i\tilde{\xi}\ln\frac{(\mathbf{q} - \mathbf{k})^2 + \kappa_1^2}{q^2 - k^2 + \kappa_1^2 - 2ik\kappa_1}\right\}. \quad (120)$$

The first term in (120) is similar to the regularized three-dimensional δ-function [cf. (85)], and the second term, to the contribution of the scattering amplitude with the energy denominator in (85). Because of regularization and the special features of Coulomb interaction, we must clarify the meaning of (119) and (120) in (87). For this purpose, let us turn to (65), which is valid for exact wavefunctions in a short-range potential. The point is that the integral of the Coulomb potential with functions of type (117) in the right-hand side of (65) also requires regularization. For instance, for a

repulsive field of *p–p* interactions, we have the exact equation [36, Appendix f]

$$\lambda e^2\int\frac{\exp(-\kappa r)}{r}\Psi_k^*(\mathbf{r})\Psi_k(\mathbf{r})d\mathbf{r} = \frac{4\pi e^2\lambda}{(2\pi)^3\kappa^2} \times \exp(-\lambda z_p)\frac{\pi\tilde{\xi}}{\sinh\pi\tilde{\xi}}F\left(i\tilde{\xi}, -i\tilde{\xi}; 1; \frac{y}{y+\eta_p^2/4}\right). \quad (121)$$

Here,

$$\tilde{\xi} = \frac{\lambda\alpha_p}{2\sqrt{y}}, \quad \alpha_p = \sqrt{\frac{m_p e^4}{\hbar^2 T}}, \quad z_p = \frac{\alpha_p}{\sqrt{y}}\arctan\left(\frac{\eta_p}{2\sqrt{y}}\right),$$

$$y = \frac{\hbar^2 k^2}{m_p T}, \quad \eta_p^2 = \frac{\hbar^2\kappa^2}{m_p T},$$

and $F(\alpha, \beta; \gamma; z)$ is the hypergeometric function.

Similar equations can be obtained for *e–e* interaction with the replacement $m_p \longrightarrow m_e$.

The left-hand side of (65) should be calculated using the $\tilde{\Psi}_k(\mathbf{q})$ functions regularized according to (118). Substituting (118) into

$$\tilde{J}_q = \int\frac{d\mathbf{q}}{(2\pi)^3}\tilde{\Psi}_k^*(\mathbf{q})(\varepsilon_k - \varepsilon_q)\tilde{\Psi}_k(\mathbf{q}) \quad (122)$$

(for instance, the first multiplier $\tilde{\Psi}_k^*(\mathbf{q})$ with the κ_1 parameter and the second one, $\tilde{\Psi}_k(\mathbf{q})$, with the regularization parameter κ_2), using the Schrödinger equation, and integrating in $d\mathbf{q}$ yields

$$\tilde{J}_q = \int d\mathbf{r}\exp(-(\kappa_1+\kappa_2)r)V(\mathbf{r})\Psi_k^*(\mathbf{r})\Psi_k(\mathbf{r}) + \frac{\hbar^2\kappa_1^2}{2\mu}\int d\mathbf{r}\exp(-(\kappa_1+\kappa_2)r)\Psi_k^*(\mathbf{r})\Psi_k(\mathbf{r}) - 2\frac{\hbar^2\kappa_1}{2\mu}\int d\mathbf{r}\frac{\exp(-(\kappa_1+\kappa_2)r)}{r}\Psi_k^*(\mathbf{r})\Psi_k(\mathbf{r}) - 2\frac{\hbar^2\kappa_1}{2\mu}\int d\mathbf{r}(\nabla\Psi_k^*(\mathbf{r})\cdot\mathbf{r})\frac{\exp(-(\kappa_1+\kappa_2)r)}{r}\Psi_k(\mathbf{r}). \quad (123)$$

A comparison of (122) and (123) with (121) leads us to assume [taking into account that (122) is real] that $\kappa_1 = \kappa_2 = \kappa/2$. It could be shown that the last three terms (counterterms) remove the square of the modulus of the first term in (120), like the square of the three-dimensional δ-function was removed in (85) and (86) to obtain the coefficient 1 in place of the doubled real part of the product of two terms in braces in (120). Taking into account the subtraction in the multiplier $(\exp(-\beta\varepsilon_k) - \exp(-\beta\varepsilon_q))$ in (87), which removes the contribution of the $f(\mathbf{k}, \mathbf{k})$-type term in (86), we find that Eq. (87) for the continuous spectrum (the spectrum of the scattering states, denoted by *SS*) should be written in the form (for instance, for *e–p* interaction)

$$\frac{\Delta\Omega_{ep}^{SS}}{V} = -16\pi^2 e^4\frac{2\mu}{\hbar^2}\zeta_p\zeta_e\lambda_{ep}^3\int_0^1 d\lambda\lambda\iint\frac{d\mathbf{q}d\mathbf{k}}{(2\pi)^6}\frac{\pi\tilde{\xi}_{ep}}{\sinh\pi\tilde{\xi}_{ep}} \times \exp(\pi\tilde{\xi}_{ep})(\exp(-\beta\varepsilon_k) - \exp(-\beta\varepsilon_q)) \times \frac{\exp(-2\tilde{\xi}_{ep}\mathrm{Im}\ln(q^2 - k_-^2))}{[(\mathbf{q}-\mathbf{k})^2 + \kappa^2/4][q^2 - k^2 - i\varepsilon]}. \quad (124)$$

Here, $k_- = k - i\kappa/2$. We can calculate the integral with respect to the angles between the vectors **q** and **k** in (124) and, after passage to the variables

$$x = \frac{\hbar^2 q^2}{2\mu T}, \quad y = \frac{\hbar^2 k^2}{2\mu T},$$

$$\eta_{ep}^2 = \frac{\hbar^2\kappa^2}{2\mu T}, \quad \tilde{\eta}_{ep} = \frac{\eta_{ep}}{2},$$

obtain

$$\frac{\Delta\Omega_{ep}^{SS}}{V} = -\frac{\lambda_{ep}}{\pi}\zeta_e\zeta_p T\left(\frac{e^2}{T}\right)^2 \times\int_0^1 d\lambda\lambda\int_0^\infty\int_0^\infty dxdy\frac{\pi\tilde{\xi}_{ep}}{\sinh\pi\tilde{\xi}_{ep}}\exp(\pi\tilde{\xi}_{ep})\frac{e^{-y}-e^{-x}}{x-y} \times\left[\frac{1}{(\sqrt{x}-\sqrt{y})^2 + \tilde{\eta}_{ep}^2} - \frac{1}{(\sqrt{x}+\sqrt{y})^2 + \tilde{\eta}_{ep}^2}\right] \times\exp\{-2\tilde{\xi}_{ep}\mathrm{Im}\ln(x - y + \tilde{\eta}_{ep}^2 + i2\sqrt{y\tilde{\eta}_{ep}^2})\}, \quad (125)$$

$$\tilde{\xi}_{ep} = \frac{\lambda\alpha_{ep}}{\sqrt{y}}, \quad \alpha_{ep} = \sqrt{\frac{\mathrm{Ry}}{T}}.$$

Let us also write the equation for the counterterm to (125) (the contribution of two crosspieces in the ladder taken into account in the summation of the series of ring diagrams). The isolation of the contribution of *e–p* interaction in (44) and the replacement of the last multiplier in the denominator, $4\pi e^2\lambda\Pi^R$, by κ^2, where κ is the regularization parameter [see (11), (118), (121)],

which should be turned to zero at the end of calculations, yields

$$\frac{\Delta\Omega_{ep}^{(2)}}{V} = -8\sqrt{\pi}\zeta_e\zeta_p\frac{e^2}{\kappa^2}\sqrt{X}\int_0^\infty e^{-y}\arctan\left(\frac{1}{2}\sqrt{\frac{\eta_{ep}^2}{y}}\right)dy$$
$$= -8\sqrt{\pi}\zeta_e\zeta_p\frac{e^2}{\kappa^2}\sqrt{X}J. \quad (126)$$

Here, J is the integral in y,

$$J = \frac{\pi}{2}\left[1 - \exp\frac{\eta_{ep}^2}{4}\left[1 - \Phi\left(\frac{\eta_{ep}}{2}\right)\right]\right], \quad (127)$$

where $\Phi(x)$ is the probability integral [38],

$$\Phi(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt. \quad (128)$$

Asymptotically as $\kappa \longrightarrow 0$, (126) and (127) give

$$\frac{\Delta\Omega_{ep}^{(2)}}{V} = -2\pi\zeta_e\zeta_p T\frac{e^4}{T^2\kappa}\left(1 - \frac{\sqrt{\pi}}{4}\kappa\lambda_{ep}\right),$$
$$\lambda_{ep} = \frac{\hbar}{\sqrt{2\mu T}}. \quad (129)$$

In order to debug the procedure for the three-dimensional integration of (125) and test the results, it is simplest to begin with calculations of the contributions of p–p and e–e interactions, because the completeness theorem for Coulomb wavefunctions in a repulsive field only includes scattering states and, taking into account (121), we obtain the identity, for instance, for p–p interaction [also see (125)],

$$\frac{\Delta\Omega_{pp}^{SS}}{V} = -\frac{\aleph_{pp}}{2\pi}\zeta_p^2 T\left(\frac{e^2}{T}\right)^2\int_0^1 d\lambda\lambda$$
$$\times\int_0^\infty\int_0^\infty dxdy\frac{\pi\tilde{\xi}_{pp}}{\sinh\pi\tilde{\xi}_{pp}}\exp(-\pi\tilde{\xi}_{pp})\frac{e^{-y} - e^{-x}}{x - y}$$
$$\times\left[\frac{1}{(\sqrt{x} - \sqrt{y})^2 + \tilde{\eta}_{pp}^2} - \frac{1}{(\sqrt{x} + \sqrt{y})^2 + \tilde{\eta}_{pp}^2}\right] \quad (130)$$
$$\times\exp\{2\tilde{\xi}_{pp}\mathrm{Im}\ln(x - y + \tilde{\eta}_{pp}^2 + i2\sqrt{y\tilde{\eta}_{pp}^2})\}$$
$$= \frac{\Delta\Omega_{pp}^{(H)}}{V} - \frac{\Delta\Omega_{pp}^{(1)}}{V}.$$

In this equation,

$$\aleph_{pp} = \sqrt{\frac{4\pi\hbar^2}{m_p T}}, \quad \tilde{\xi}_{pp} = \frac{\lambda\alpha_p}{2\sqrt{y}},$$
$$\alpha_p = \sqrt{\frac{m_p e^4}{\hbar^2 T}}, \quad y = \frac{\hbar^2 k^2}{m_p T}, \quad x = \frac{\hbar^2 q^2}{m_p T},$$
$$\eta_{pp}^2 = \frac{\hbar^2\kappa^2}{m_p T}, \quad \tilde{\eta}_{pp} = \frac{\eta_{pp}}{2}.$$

For $\Delta\Omega_{pp}^{(H)}$ from (121) [the index "(H)" denotes the hypergeometric function present in (121)], we obtain

$$\frac{\Delta\Omega_{pp}^{(H)}}{V} = 4\sqrt{\pi}\zeta_p^2\frac{e^2}{\kappa^2}\int_0^1 d\lambda\int_0^\infty dy\sqrt{y}e^{-y}e^{-\lambda z_p}\Phi. \quad (131)$$

Here,

$$z_p = \frac{\alpha_p}{\sqrt{y}}\arctan\left(\frac{\eta_{pp}}{2\sqrt{y}}\right), \quad (132)$$

$$\Phi = F\left(i\tilde{\xi}_{pp}, -i\tilde{\xi}_{pp}; 1; \frac{y}{y + \eta_{pp}^2/4}\right)\frac{\pi\tilde{\xi}_{pp}}{\sinh\pi\tilde{\xi}_{pp}}. \quad (133)$$

The $\Delta\Omega_{pp}^{(1)}$ value, which corresponds to the subtraction of the diagram with one crosspiece [of the type of the subtraction of the ~$\exp(-\beta\varepsilon_q)$ term in parentheses in (87)], in the ladder diagram is given by

$$\frac{\Delta\Omega_{pp}^{(1)}}{V} = \frac{2\pi e^2\zeta_p^2}{\kappa^2}. \quad (134)$$

When summation is performed, all contributions of p–p, e–e, and e–p interactions vanish by virtue of (9) and (13). To remove the divergence of $\Delta\Omega_{pp}^{SS}$ and $\Delta\Omega_{pp}^{(H)} - \Delta\Omega_{pp}^{(1)}$, we must subtract the contribution of two crosspieces from (130)–(134) [as with (126)–(129)],

$$\frac{\Delta\Omega_{pp}^{(2)}}{V} = -4\sqrt{\pi}\zeta_p^2\frac{e^2}{\kappa^2}\int_0^\infty\sqrt{y}\exp(-y)\frac{z_p}{2}dy$$
$$\longrightarrow -\pi\zeta_p^2 T\frac{e^4}{T^2\kappa}\left(1 - \frac{\aleph_{pp}\kappa}{8}\right) \text{ as } \kappa \longrightarrow 0. \quad (135)$$

Numerical integration was performed using library subroutines for calculating integrals over a finite one-dimensional interval based on the Gauss–Kronrod quadrature formulas (the number of nodes was 10 and 21). The number of points was selected automatically, and an adaptive procedure was used for mesh refinement close to sharp peaks of the integrand to attain the required accuracy. We also developed a nested procedure for calculating multidimensional integrals by recursion with respect to integral dimension.

At each nesting level, the integrand was represented as a nested integral and calculated by a similar recursion using a separate copy of the library subroutines for one-dimensional integration.

One-dimensional integration was performed with setting the required relative accuracy. In multidimensional integration, the relative accuracy of calculating the integrand (that is, each integral "inner" with respect to the embedment rank) at each given point was set at a value inversely proportional to the magnitude of its contribution to the integral "external" in its rank.

The three-dimensional integrals in (125) and (130) were calculated in the variables

$$u = \sqrt{x} + \sqrt{y}, \quad v = \sqrt{x} - \sqrt{y}, \quad \lambda$$

in the following order:

$$\int_0^\infty du \int_{-u}^{u} dv \int_0^1 d\lambda.$$

The integration in v required the interval to be additionally partitioned into subintervals whose ends coincided with local maxima. The first and second crosspieces (which increased as $\kappa \longrightarrow 0$ proportionally to κ^{-2} and κ^{-1}, respectively) were calculated using analytic equations similar to (134), (126), and (127).

Note that the use of library subroutines for integration over semi-infinite intervals to calculate the second crosspiece led to errors random in κ at small κ values, which were outside the required relative accuracy. When constants asymptotic in κ were calculated for differences of the type $\Delta\Omega_{ep}^{(2)} - \Delta\Omega_{ep}^{SS}$, the error was comparable to the result. For this reason, integrals over semi-infinite intervals were calculated iteratively, as the limiting values of the integrals over segments when the upper integration limit was successively doubled. This variant always gave small stable errors in approximations to the second crosspiece at all κ values and was used for integration with infinite limits.

At the first stage, we obtained satisfactory accuracy in reproducing the asymptotic behavior, linear in $\ln\eta_f$ ($\eta_f = \kappa/\kappa_D$, $\eta_f \longrightarrow 0$), for the equations

$$\begin{aligned} \mathcal{J}_{aa} &= e^{-6}T^2V^{-1}\zeta_a^{-2} \\ &\times (\Delta\Omega_{aa}^{(H)} - \Delta\Omega_{aa}^{(1)} - \Delta\Omega_{aa}^{(2)}), \\ a &= e, p, \\ \mathcal{J}_{ep} &= e^{-6}T^2V^{-1}\zeta_e^{-1}\zeta_p^{-1}(\Delta\Omega_{ep}^{(2)} - \Delta\Omega_{ep}^{SS}). \end{aligned} \tag{136}$$

The right-hand side of (136) turned asymptotically to

$$-A - k_{ab}\frac{\pi}{3}\ln\eta_f,$$

where

$$k_{ee} = k_{pp} = 1, \quad k_{ep} = 2$$

(see [20] for *e–e* and *p–p* interactions). Subsequent calculations of $A = A(\rho, T)$ with the highest accuracy possible for obtaining a suitable analytic approximation requires the accuracy of calculating $\Delta\Omega_{aa}^{(H)}/V$ and $\Delta\Omega_{ep}^{SS}/V$ to be increased as required by a decrease in $\eta_f \longrightarrow 0$.

Next, the construction of equations similar to [13, (6.1)–(6.4)] and calculations at several (ρ, T) points along the Sun radius gave the asymptotic equation

$$\begin{aligned} &\mathcal{J}_{ab}(\kappa, T) \xrightarrow[\kappa \to 0]{} -A_{ab}(T) \\ &- k_{ab}\left(\frac{\pi}{3}\ln(\kappa\lambda_{ab}) - \frac{\pi}{6}(1 - C)\right), \\ &ab = ep, ee, pp. \end{aligned} \tag{137}$$

The value

$$\begin{aligned} &\tilde{A}_{ab}(T, \kappa) \\ &= -\left[\mathcal{J}_{ab}(\kappa, T) + k_{ab}\left(\frac{\pi}{3}\ln(\kappa\lambda_{ab}) - \frac{\pi}{6}(1 - C)\right)\right] \end{aligned} \tag{138}$$

converges with respect to κ to the limiting value

$$A_{ab}(T) = \lim_{\kappa \to 0} \tilde{A}_{ab}(T, \kappa),$$

and the rate of convergence (absolute error)

$$\varepsilon(T, \kappa) = \left|\tilde{A}_{ab}(T, \kappa) - A_{ab}(T)\right| \tag{139}$$

is proportional to κ^2 for *e–e* and *p–p* interactions and $\sim\kappa^{1.14}$ for *e–p* interaction. The smallest $\tilde{\eta}_{ep}^2$ value at which we were able to calculate integral (125) and estimate A_{ab} was on the order of 10^{-12}–10^{-13}.

The limiting A_{ab} values ($ab = ep$, ee, and pp) are shown in Fig. 2 versus the $|\alpha_{ab}|$ parameter ($\alpha_{ee} = -\alpha_e/2$, $\alpha_{pp} = -\alpha_p/2$). It turned out that these values fairly closely (indistinguishably on a linear scale) coincide with the analytic equations

$$\begin{aligned} B_{ep} &= \chi_{ep}^3 T^3 e^{-6}(\Sigma_{SRM}^{BS} - \Sigma^{\text{tot}}), \\ \Sigma^{\text{tot}} &= \Sigma_Q(\alpha_{ep}), \end{aligned} \tag{140}$$

$$\begin{aligned} B_{ee} &= \frac{1}{2}\chi_{ee}^3 T^3 e^{-6}\Sigma_{ee}^{SS}, \\ \Sigma_{ee}^{SS} &= \Sigma_Q\left(-\frac{\alpha_e}{2}\right), \quad \alpha_e = \sqrt{\frac{m_e e^4}{\hbar^2 T}}, \end{aligned} \tag{141}$$

$$\begin{aligned} B_{pp} &= \frac{1}{2}\chi_{pp}^3 T^3 e^{-6}\Sigma_{pp}^{SS}, \\ \Sigma_{pp}^{SS} &= \Sigma_Q\left(-\frac{\alpha_p}{2}\right), \quad \alpha_p = \sqrt{\frac{m_p e^4}{\hbar^2 T}}, \end{aligned} \tag{142}$$

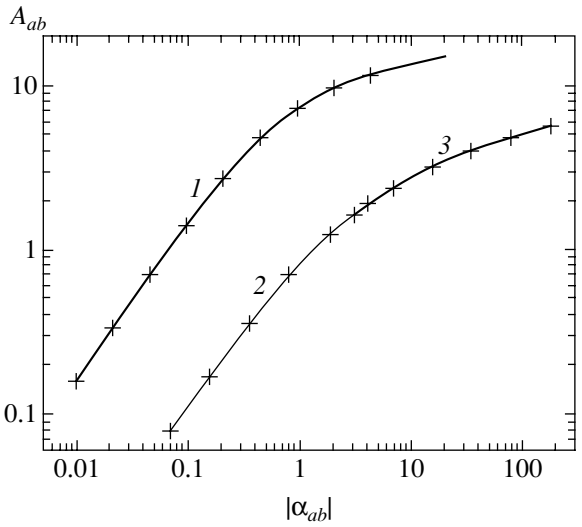


Fig. 2. Calculated A_{ab} values versus the $|\alpha_{ab}|$ parameter; *ab*: (*1*) *e–p*, (*2*) *e–e*, and (*3*) *p–p* interactions. The corresponding analytic B_{ab} values calculated at several points are marked by "+."

where the $\Sigma_Q(\alpha)$ function is represented by the series [see Section 6, Eq. (184)]

$$\Sigma_Q(\alpha) = \frac{1}{2}\sum_{n=4}^{\infty}\frac{\zeta(n-2)}{\Gamma\left(\frac{n}{2}+1\right)}\alpha^n. \tag{143}$$

At large $-\alpha \gg 1$, the asymptotic behavior of (143) is described by the equation

$$\Sigma_Q^0(\alpha) = -\left(\ln|2\alpha| + \frac{3C}{2} - \frac{4}{3}\right)\frac{2\alpha^3}{3\sqrt{\pi}} + o(|\alpha|^3), \quad \alpha \longrightarrow -\infty. \tag{144}$$

The largest deviation

$$\max_{\alpha}|A_{ab} - B_{ab}|$$

along the solar curve is close to 2.5×10^{-4} for *e–e*, 10^{-3} for *p–p*, and 10^{-2} for *e–p* interaction. The larger error in the *e–p* asymptotic is explained by the lower rate of convergence and the necessity of calculating three-dimensional integrals. For *e–e* and *p–p* interactions, we used two-dimensional integration of (131)–(133) after verifying the validity of (130). Note that the limiting relative accuracy of calculations was 2×10^{-14} for two-dimensional and 10^{-10} for three-dimensional integrals.

The conclusion can be drawn that numerical integration with an accuracy to the excessive term with ln3 in [13] substantiates the analytic representation of the Born series given in [13].

Note that (131) can be integrated analytically for comparison with and verification of the equations obtained in [13]. First, let us write an equation for the hypergeometric function [40] whose argument γ equals 1,

$$F(i\tilde{\xi}, -i\tilde{\xi}; 1; z) = \frac{\sinh\pi\tilde{\xi}}{\pi\tilde{\xi}}\left\{1 + \frac{(\kappa\tilde{\xi}/2)^2}{k^2 + \kappa^2/4}\right. \times\left.\left[\ln\frac{\kappa^2/4}{k^2+\kappa^2/4} + 2C - 1 + 2\mathrm{Re}\,\psi(1+i\tilde{\xi})\right]\right\} + O(\kappa^4). \tag{145}$$

Here, $C = 0.5772\ldots$ is the Euler–Maclaurin constant [38] and $\psi(x) = d\ln\Gamma(x)/dx$ is the ψ-function. The $\psi(1 + i\tilde{\xi})$ value can be expanded into a series in powers of $\tilde{\xi}$ [38],

$$2\mathrm{Re}\,\psi(1+i\tilde{\xi}) = -2C - 2\sum_{k=1}^{\infty}\tilde{\xi}^{2k}\zeta(2k+1), \quad \tilde{\xi} = \frac{\lambda\alpha_p}{2\sqrt{y}}. \tag{146}$$

Substituting (146) into (131) immediately shows that series (146) terms give divergent results when integrated in *y*. We can ignore the term $\kappa^2/4$ compared with k^2 because of the passage to the limit $\kappa \longrightarrow 0$ in the coefficient of the sum in square brackets and under the logarithm sign.

Let us use the following procedure for the summation of a series of diverging integrals, which turns out to be rigorous although slightly unusual. Consider a model example, for which all procedures are similar to those that arise when (131) is handled with the use of (145) and (146). We use the integral representation of the ψ-function [38]

$$J_{\text{model}} = \mathrm{Re}\int_0^{\infty}\psi\left(1 \pm i\frac{a}{\sqrt{y}}\right)e^{-y}\frac{dy}{\sqrt{y}} = -C\sqrt{\pi} + \sqrt{\pi}\int_0^{\infty}dt\frac{e^{-t}}{1-e^{-t}} - \int_0^{\infty}dt\frac{e^{-t}}{1-e^{-t}}\int_0^{\infty}\mathrm{Re}\exp\left(\mp i\frac{at}{\sqrt{y}}\right)\frac{e^{-y}}{\sqrt{y}}dy. \tag{147}$$

Here, *a* is a real number. Let us introduce the denotation $b = |a|t$ and pass to the new variable $x = y^{-1/2}$ in the inner

integration in the last term in (147). This yields [41, p. 454, (2.5.38.1–3)]

$$\mathrm{Re}\int_0^\infty \exp\left(\mp i\frac{at}{\sqrt{y}}\right)\frac{e^{-y}}{\sqrt{y}}dy = 2\int_0^\infty \cos bx \exp(-1/x^2)\frac{dx}{x^2} = \pi\sum_{k=0}^\infty \frac{(-|b|)^k}{k!\Gamma\left(\frac{k+1}{2}\right)}. \quad (148)$$

Using the relation [38]

$$\int_0^\infty dt\frac{e^{-t}}{1-e^{-t}}t^k = \Gamma(k+1)\zeta(k+1), \quad (149)$$

and (148) in (147), we obtain

$$J_{\mathrm{model}} = -C\sqrt{\pi} - \pi\sum_{k=1}^\infty \frac{(-1)^k|a|^k}{\Gamma\left(\frac{k+1}{2}\right)}\zeta(k+1). \quad (150)$$

This is a rigorous result. On the other hand, we can expand $\psi(1 \pm ia/\sqrt{y})$ in powers of a as in (146),

$$\mathrm{Re}\,\psi\left(1 \pm i\frac{a}{\sqrt{y}}\right) = -C - \sum_{m=1}^\infty (-1)^m\zeta(2m+1)\frac{a^{2m}}{y^m}, \quad (151)$$

and integrate this series term by term using the definition of J_{model} [Eq. (147)],

$$J_{\mathrm{model}} = -C\sqrt{\pi} - \pi\sum_{m=1}^\infty (-1)^{2m}\frac{a^{2m}\zeta(2m+1)}{\Gamma\left(m+\frac{1}{2}\right)}. \quad (152)$$

The diverging integrals in the right-hand side of (152) were regularized using analytic continuations (relations for the gamma function [38]),

$$\int_0^\infty \frac{y^{1/2}}{y^{m+1}}e^{-y}dy = \Gamma\left(-m+\frac{1}{2}\right) = \frac{\pi}{\cos\pi m\,\Gamma\left(m+\frac{1}{2}\right)} = \frac{(-1)^m\pi}{\Gamma\left(m+\frac{1}{2}\right)}. \quad (153)$$

A comparison of (150) and (152) shows that (150) contains all integer powers of a, whereas (152), even powers only. Passing to the new summation variable

$$2m = k \quad (m = k/2) \quad (154)$$

and performing the summation over all k from 1 rather than over even k from 2 only, as we ought to (this is what appears unusual), we obtain from (152)–(154)

$$J_{\mathrm{model}} = -C\sqrt{\pi} - \pi\sum_{k=1}^\infty (-1)^k\frac{|a|^k\zeta(k+1)}{\Gamma\left(\frac{k+1}{2}\right)},$$

which exactly coincides with the rigorous solution [Eq. (150)].

After the subtraction of (134) and (135) from (131), we can apply regularization procedure (153) and partition the summation variable (154) to obtain from (147)

$$\frac{\delta\Omega_{pp}^L}{V} = \frac{\delta\Omega_{pp}^{cl}}{V} + \frac{\delta\Omega_{pp}^q}{V} \quad (155)$$

[as in the derivation of (150)]. The first term on the right-hand side of (155) corresponds to the "classical" limit represented by the logarithmic contribution to the second virial coefficient (here, $\kappa \longrightarrow 0$),

$$\frac{\delta\Omega_{pp}^{cl}}{V} = -\zeta_p^2 T\left(\frac{e^2}{T}\right)^3\left\{\frac{\pi}{3}\ln\kappa\lambda_p - \frac{\pi}{6}(1-C)\right\}, \quad \lambda_p = \frac{\hbar}{\sqrt{m_p T}}. \quad (156)$$

In [13], the corresponding equation contains an unnecessary term with ln3 in parentheses, $(-C - 2\ln 3 + 1)$. This term arises when the classical representation for the second virial coefficient (B_{cl}) is used "literally" with the expansion in powers of U/T up to the third order and the replacement of

$$U(r) = e^2/r$$

by the screened interaction potential value

$$U(r) = (e^2/r)\exp(-\kappa_D r).$$

In [13], κ in (156) is understood to be κ_D [28]. We have

$$B_{cl} = \frac{T}{2}\int_0^\infty\left(\exp\left(-\frac{U}{T}\right)-1\right)dU \longrightarrow -\frac{T}{2}\frac{1}{6}\int_0^\infty\left(\frac{e^2}{T}\right)^3\frac{1}{r^3}\exp(-3\kappa_D r)4\pi r^2 dr \longrightarrow -T\left(\frac{e^2}{T}\right)^3\frac{\pi}{3}\ln 3\kappa_D(r_{\min}). \quad (157)$$

The lower limit is set at $\max(e^2/T, \lambda_p)$. The regularization procedure [Eq. (121)] contains a linear exponent, and ln3 does not appear. Note that (156) is obtained using analytic continuations and the equations [38]

$$\int_0^\infty y^{-\frac{1}{2}-1} e^{-y} \ln\Lambda dy = \ln\Lambda \cdot \Gamma\left(-\frac{1}{2}\right) = -2\sqrt{\pi}\ln\Lambda. \quad (158)$$

Here, $\Lambda = (\kappa\lambda_p/2)^2$ and

$$\int_0^\infty y^{-\frac{1}{2}-1} e^{-y} \ln y dy = \Gamma\left(-\frac{1}{2}\right)\psi\left(-\frac{1}{2}\right) = -2\sqrt{\pi}(-C + 2 - 2\ln 2). \quad (159)$$

We will show that the sum of the contributions $\delta\Omega_{pp}^L$, $\delta\Omega_{ee}^L$, and $\delta\Omega_{ep}^L$ leads to the disappearance of the regularization parameter $\kappa \longrightarrow 0$ in the final solution for a hydrogen plasma.

The second, quantum, term in (155) is given by a series in powers of the Born parameter (which is on the order of $e^2/\hbar v_T$), that is, the $\alpha_p/2$ parameter in our problem. This series exactly coincides with that given in [13],

$$\frac{\delta\Omega_{pp}^q}{V} = -2\pi\lambda_p^3\zeta_p^2 T\sum_{n=4}^\infty \frac{\zeta(n-2)}{\Gamma\left(\frac{n}{2}+1\right)}\left(-\frac{\alpha_p}{2}\right)^n. \quad (160)$$

We stress that this result is also substantiated by a detailed comparison with numerical integration in both representations (130) and (131) after the subtraction of $\delta\Omega_{pp}^{(1)}$ from (131) and $\delta\Omega_{pp}^{(2)}$ from both (130) and (131). Solutions (156) and (160) with the replacement $m_p \longrightarrow m_e$ are also valid for *e*–*e* interaction.

Identity (130) does not hold for attractive *e*–*p* interaction because the use of the completeness theorem for Coulomb functions that leads to (134) requires the inclusion of discrete and continuous states, whereas continuous states are only present in (130).

The numerical integration results show that the three-dimensional integral is approximated by an analytic equation whose meaning is discussed in Section 6, namely,

$$\frac{\delta\Omega_{ep}^L}{V} = \frac{\delta\Omega_{ep}^{cl}}{V} + \frac{\delta\Omega_{ep}^q}{V}. \quad (161)$$

Here, $\delta\Omega_{ep}^{cl}$ is the logarithmic contribution [cf. (156)]

$$\frac{\delta\Omega_{ep}^{cl}}{V} = \zeta_e\zeta_p T\left(\frac{e^2}{T}\right)^3 2\left\{\frac{\pi}{3}\ln\kappa\lambda_{ep} - \frac{\pi}{6}(1-C)\right\}, \quad \lambda_{ep} = \frac{\hbar}{\sqrt{2\mu T}}. \quad (162)$$

Summing the equations of type (156) for *p*–*p* and *e*–*e* interactions and (162) yields

$$\frac{\delta\Omega^{cl}}{V} = \zeta_e^2 T\left(\frac{e^2}{T}\right)^3\frac{\pi}{6}\ln\frac{m_p}{4m_e} \quad (163)$$

(it is taken into account that $\zeta_e = \zeta_p$). This result does not contain the κ parameter.

The solution for $\delta\Omega_{ep}^q$ can be written in the form [cf. (105), (108)]

$$\frac{\delta\Omega_{ep}^q}{V} = -\zeta_e\zeta_p T\lambda_{ep}^3\Sigma^{SS}, \quad (164)$$

$$\Sigma^{SS} = \sum_{n=4}^\infty \frac{\zeta(n-2)}{\Gamma\left(\frac{n}{2}+1\right)}\left(\frac{1}{2} - (-1)^n(n-2)^2\right)\alpha_{ep}^n - \sum_{k=1}^\infty \frac{\zeta(2k+1)}{\Gamma\left(k+\frac{5}{2}\right)}\alpha_{ep}^{2k+3}, \quad \alpha_{ep} = \sqrt{\beta\mathrm{Ry}}. \quad (165)$$

The equation for exchange interaction (mainly, *e*–*e*) can be obtained by analogy with (121) using (81) and (82).

Equations of type (117) can be used to obtain J_{exch} for a repulsive field in the form

$$J_{\mathrm{exch}} = \int d\mathbf{r}\Psi_k^*(\mathbf{r})V(\mathbf{r})\Psi_k(-\mathbf{r}) = \frac{1}{(2\pi)^3}\frac{\pi\tilde{\xi}_{ee}}{\sinh\pi\tilde{\xi}_{ee}}\lambda e^2\pi, \quad \tilde{\xi}_{ee} = \frac{\lambda}{\sqrt{2y}}\sqrt{\frac{m_e e^4}{2\hbar^2 T}}, \quad y = \frac{\hbar^2 k^2}{m_e T}. \quad (166)$$

Eventually, the exchange contribution $\delta\Omega_{ee}^{\mathrm{exch}}$ is written as

$$\frac{\delta\Omega_{ee}^{\mathrm{exch}}}{V} = -\frac{\zeta_e^2}{16\hbar^2}\lambda_{ee}^3 m_e e^4\int_0^1 \lambda d\lambda\int_0^\infty \frac{dk}{k}\frac{\exp\left(-\frac{\hbar^2 k^2}{m_e T}\right)}{\sinh\frac{\pi\lambda m_e e^2}{2\hbar^2 k}}. \quad (167)$$

The integrals in (167) are calculated numerically. We can also obtain an explicit analytic result using a technique similar to that applied to derive Eqs. (147)–(150) for J_{model} and the expansion

$$\frac{1}{\sinh x} = \frac{1}{x} - \sum_{k=1}^{\infty} \frac{2(2^{2k-1}-1)}{(2k)!} B_{2k} x^{2k-1}. \quad (168)$$

With Bernoulli numbers given by [38]

$$B_{2k} = (-1)^{k-1} \frac{2(2k)!}{(2\pi)^{2k}} \zeta(2k), \quad B_{2k+1} = 0, \quad (169)$$

(167) yields the initially converging result

$$\frac{\delta\Omega_{ee}^{\text{exch}}}{V} = \frac{\zeta_e^2 T}{\sqrt{\pi}} \aleph_{ee}^3 E(\alpha_e), \quad \aleph_{ee} = \sqrt{\frac{4\pi\hbar^2}{m_e T}}, \quad (170)$$

where $E(\alpha_e)$ is given by

$$E(\alpha_e) = -\frac{\alpha_e}{4} + \frac{\sqrt{\pi}}{4} \ln 2 \left(-\frac{\alpha_e}{2}\right)^2 + \frac{\pi^2}{72}\left(-\frac{\alpha_e}{2}\right)^3 + \sum_{n=4}^{\infty} \frac{\sqrt{\pi}(1-2^{2-n})}{\Gamma\left(\frac{n}{2}+1\right)} \zeta(n-1)\left(-\frac{\alpha_e}{2}\right)^n, \quad (171)$$

$$\alpha_e = \sqrt{\frac{m_e e^4}{\hbar^2 T}}.$$

Equation (171) exactly coincides with that given in [13] for the exchange contribution and is substantiated by numerical integration. The first Born series (171) term gives the contribution described above by (15). Note that, even for the conditions close to the center of the Sun, the Born series for *p*–*p* interaction [Eq. (160)] converges poorly because $\alpha_p \gg 1$.

In this limit, expansion (146) can be replaced by the asymptotic equation for the ψ-function [38] ($\tilde{\xi} \gg 1$),

$$\mathrm{Re}\psi(1 + i\tilde{\xi}) = \ln\tilde{\xi}. \quad (172)$$

Substituting (172) into (131) and using (145) and the method for summing diverging integrals [similar to (147)–(150)], we obtain

$$\frac{\delta\Omega_{pp}^{L}}{V} \approx -\zeta_p^2 T \left(\frac{e^2}{T}\right)^3 \frac{\pi}{3}\left\{\ln\Gamma + 2C - \frac{11}{6}\right\}, \quad (173)$$

$$\Gamma = e^2\kappa/T, \quad \kappa \longrightarrow 0$$

in the limit $\alpha_p \gg 1$.

In this limit, it follows from (173) ($\alpha_p \gg 1$) and equations of types (163) (for *e*–*e* interaction at $\alpha_e \ll 1$) and (162) ($\alpha_{ep} \ll 1$) that ($\zeta_e = \zeta_p$)

$$\frac{\delta\Omega^{L}}{V} = \zeta_e^2 T \left(\frac{e^2}{T}\right)^3 \frac{\pi}{6} \times \left\{\ln\frac{m_p}{2\mu_{ep}} + 2\ln\frac{\lambda_{ep} T}{e^2} - 3C + \frac{8}{3}\right\}. \quad (174)$$

Equation (174) also does not contain κ ($\lambda_{ep}T/e^2 \gg 1$ according to the condition $\alpha_{ep} \ll 1$). Note in addition that, in the limit $\alpha_p \gg 1$, $\alpha_e \gg 1$, and $\alpha_{ep} \gg 1$, the contribution of all logarithmic corrections, including the constants outside the logarithm sign, is cancelled for a hydrogen plasma. For a plasma comprising electrons and nuclei with charge $z \gg 1$ and atomic weight *A*, the derivation of the equation for the second virial coefficient performed ignoring the contribution of the constants outside the logarithm sign ($\ln z \gg 1$) yields the following results:

(1) In the limit

$$\alpha_p^z \ll 1, \quad \alpha_e \ll 1, \quad \alpha_{ep}^z \ll 1,$$

where

$$\alpha_p^z = z^2 \sqrt{\frac{m_p e^4 A}{\hbar^2 T}}, \quad \alpha_{ep}^z = z\sqrt{\beta \mathrm{Ry}}$$

(this limit presupposes anomalously high temperatures $T \gtrsim 50 z^4 A$ keV), we obtain

$$\frac{\delta\Omega^{cl}}{V} = -\zeta_e^2 T \left(\frac{e^2}{T}\right)^3 \frac{\pi}{3} \times \left\{(z^2-1)^2 \ln\frac{\kappa\hbar}{\sqrt{m_e T}} - \frac{z^4}{2}\ln\frac{m_e}{m_p A} + z^2 \ln 2\right\}. \quad (175)$$

Here, the condition $\zeta_e = z\zeta_p$ is used. Equation (175) explicitly contains the κ parameter (at $z \neq 1$); it is reasonable to set this parameter equal to κ_D, $\kappa = \kappa_D$.

(2) In the limit

$$\alpha_p^z \gg 1, \quad \alpha_e \gg 1, \quad \alpha_{ep}^z \gg 1,$$

we obtain

$$\frac{\delta\Omega^{cl}}{V} = -\zeta_e^2 T \left(\frac{e^2}{T}\right)^3 \frac{\pi}{3} \times \{(z^2-1)^2 \ln\Gamma_D + 2z^2(z^2-1)\ln z\}. \quad (176)$$

Here, it is assumed that

$$\Gamma = \Gamma_D = \kappa_D e^2/T \ll 1, \quad \Gamma_D z^2 \ll 1.$$

At $z \gg 1$, (176) can be written in the form

$$\frac{\delta\Omega^{cl}}{V} = -T\zeta_e \frac{\Gamma_D^2 z^3}{12} \ln \Gamma_D z^2. \tag{177}$$

6. THE EQUATION OF STATE OF A WEAKLY NONIDEAL HYDROGEN PLASMA

Consider the summed contribution of Eqs. (6), (8), (16), (52), (57), (98), (104), (155), (156), (163), (160), (164), (165), (170), and (171) obtained above as applied to describe the equation of state of a weakly nonideal hydrogen plasma in relation to helioseismology problems.

The meaning of these equations is as follows: (6) describes the pressure of the ideal gas of protons; (8), the pressure of the ideal gas of degenerate electrons; (16), the contribution of Coulomb interaction to the pressure ($\delta P_{D-H} = -\delta\Omega_{D-H}/V$) in the Debye–Hückel approximation in a grand canonical ensemble; (52), the diffraction (quantum) corrections to the Debye–Hückel approximation; (57), corrections taking into account that particles interacting with the medium, rather than free particles, are responsible for Debye screening; (98) and (104), the converging contribution of bound states to the pressure (the contribution of "atoms"); (155) and (156), the logarithmic contribution on the order of $(e^2/T)^3$ of proton–proton interaction to the second virial coefficient; (163), the total logarithmic contribution of all interactions on the order of $(e^2/T)^3 \ln(m_p/m_e)$ in a plasma of protons and electrons; (160), (164), and (165), the contribution of scattering states to the thermodynamic characteristics of a plasma; and (170) and (171), the contribution of electron–electron exchange.

It is necessary to take into account relativistic corrections to the contribution of the free electron gas to thermodynamic functions and radiation pressure in a "transparent" plasma in the center of the Sun. Treating relativistic contributions as a correction ($T/mc^2 \lesssim 10^{-3}$), let us write the relation between the momentum and kinetic energy of the electron as

$$p = \sqrt{2mE}\left(1 + \frac{E}{4mc^2}\right). \tag{178}$$

Equation (178) can be used to determine the activity of electrons [cf. (7)]

$$n_e^0(\mu_e) \equiv \zeta_{e,r} = \frac{2}{\sqrt{\pi}\lambda_e^3}\int_0^\infty \frac{x^{1/2}\left(1 + \frac{5}{4}\frac{T}{mc^2}x\right)}{\exp(x-y)+1}dx \tag{179}$$

and the electron gas pressure [cf. (8)]

$$P_{0e,r} = \frac{4T}{3\sqrt{\pi}}\frac{2}{\lambda_e^3}\int_0^\infty \frac{x^{3/2}\left(1 + \frac{3}{4}\frac{T}{mc^2}x\right)}{\exp(x-y)+1}, \tag{180}$$

(the index "*r*" means that relativistic corrections are included). As previously,

$$x = \varepsilon/T, \quad y = \mu_e/T$$

in (179) and (180).

Corrections to the energy and other functions of electrons can be found in a similar way. To obtain the equation of state of a weakly nonideal hydrogen plasma, we must find a relation between activities (or chemical potentials) and concentration [cf. (24)] taking into account the degeneracy of electrons,

$$n_e + \sum_k n_k = \sum_k (z_k + 1) n_k = -\sum_k{}' \left(\frac{\partial(\Omega/V)}{\partial\mu_k}\right)_T. \tag{181}$$

The thermodynamic potential Ω is represented by the sum of all the contributions specified above. Here, the primed sum denotes that the differentiation in (181) should be performed taking into account (13) and equations similar to (179) and (180) for the activity and pressure of the electrons. For instance, within the framework of the simplest model of type (20), for which, in addition to the ideal gas contribution, the Debye–Hückel contribution is taken into account with the use of general equations (16) and (17) and neutrality condition (19), we obtain for the simplest case of $z_k = 1$

$$2n_e = \zeta_e + \zeta_p + \frac{\Gamma_D}{2}\left(\zeta_p + T^2\frac{\partial^2 n_e^0}{\partial\mu_e^2}\right). \tag{182}$$

In (182), n_e^0 should be differentiated with respect to the chemical potential taking into account (179). Next,

we must put

$$\zeta_{e,r} = \zeta_p = \zeta, \quad n_e = N_A\rho,$$

where N_A is Avogadro's number and ρ is the plasma density.

Note that, because a hydrogen plasma is only weakly nonideal ($\Gamma_D \ll 1$) and corrections for degeneracy and relativistic contributions are comparatively small for it under Sun conditions, the contributions to the second virial coefficient other than those included in model (20) can be described by Eqs. (52), (57), (98), (104), (155), (156), (163), (160), (164), (165), (170), and (171) obtained for a plasma with nondegenerate electrons.

Note that the analytic equations for the converging contribution of bound states [Eq. (104)] together with the final equation for the contribution of scattering states [Eqs.(164) and (165)] [strictly speaking, this equation was a guess substantiated by a comparison with the results of numerical integration according to (125) from which (126) was subtracted] and with the use of expansion (105) give

$$\frac{\delta\Omega_{ep}^q}{V} = \frac{\delta\Omega^{BS}}{V} + \frac{\delta\Omega_{ep}^{SS}}{V} = -\zeta_e\zeta_p T\varkappa_{ep}^3(\Sigma^{BS} + \Sigma^{SS}) = -\zeta_e\zeta_p T\varkappa_{ep}^3\Sigma^{\text{tot}}. \quad (183)$$

Here, Σ^{tot} is the sum of the series in powers of $\alpha_{ep} = \sqrt{\beta\text{Ry}}$, which exactly coincides with the equations given in [13, 14] and obtained from Beth–Uhlenbeck-type equations written for the total contribution of attractive states without partitioning them into bound and scattering states,

$$\Sigma^{\text{tot}} = \frac{1}{2}\sum_{n=4}^{\infty}\frac{\zeta(n-2)}{\Gamma\left(\frac{n}{2}+1\right)}\alpha_{ep}^n. \quad (184)$$

Interestingly, Eq. (184) was obtained in [13, 14] by discarding the divergent terms of order e^2, e^4, and e^6 on account of Coulomb screening. Our approach allows the exact correspondence to be obtained between (184) and the sum of converging contributions of bound states (which do not at all require taking screening into account) and scattering states calculated using three-dimensional integrals (125), from which the contribution of two crosspieces [Eq. (126)] should be subtracted; the latter were already included in obtaining the ring approximation corresponding to the Debye–Hückel approximation. At large parameter values $\alpha_{ep} \gg 1$, the sum in (184) can be estimated by replacing summation with integration in n and applying the saddle-point method,

$$\Sigma^{\text{tot}} \longrightarrow \exp(I/T),$$

which corresponds to taking into account the ground state only in the atomic partition function, as was initially done in the Saha equation. Using this formula to estimate the fraction of bound states ("atoms") in the center of the Sun gives a value on the order of 30% (an absurd result), whereas, according to the Planck–Larkin equation, this fraction is 10^{-4} in order of magnitude. If the contribution of bound states represented by the Planck–Larkin equation [with $\Sigma_{\text{P–L}}^{BS}$ given by (3) and (106)] is subtracted from (184), we can obtain an analogue of the contribution of scattering states [cf. (164), (165)],

$$(\Sigma^{SS})' = \Sigma^{\text{tot}} - \Sigma_{\text{P–L}}^{BS} = -\frac{1}{2}\sum_{n=4}^{\infty}\frac{(-1)^n\zeta(n-2)}{\Gamma\left(\frac{n}{2}+1\right)}\alpha_{ep}^n. \quad (185)$$

Equation (185) was not written in explicit form. However, comparing it with the equation for the contribution of scattering states of type (160) for p–p interaction shows that this contribution has a similar structure. [Also compare (148), where the result is independent of the sign of a].

Equations (107), (108), and (185) therefore allow us to write the relations

$$\Sigma^{SS} = (\Sigma^{SS})' + \Sigma_c, \quad (186)$$

$$\Sigma_{SRM}^{BS} = \Sigma_{\text{P–L}}^{BS} - \Sigma_c. \quad (187)$$

The total contribution of bound and scattering states [Eq. (184)] can be written in two ways,

$$\Sigma^{\text{tot}} = \Sigma_{SRM}^{BS} + \Sigma^{SS} = \Sigma_{\text{P–L}}^{BS} + (\Sigma^{SS})'. \quad (188)$$

In the sense of (188), the Plank–Larkin equation is "correct." It, however, contains contributions of bound states and the continuous spectrum, provided the contribution of scattering states is determined properly. Unlike the Planck–Larkin equation, the equation for Σ_{SRM}^{BS} describes the contribution of bound states only.

Taking into account that $\delta\Omega/V = -T\zeta_e\zeta_p\varkappa_{ep}^3\Sigma$, let us write the first several terms of (105), (106), (165),

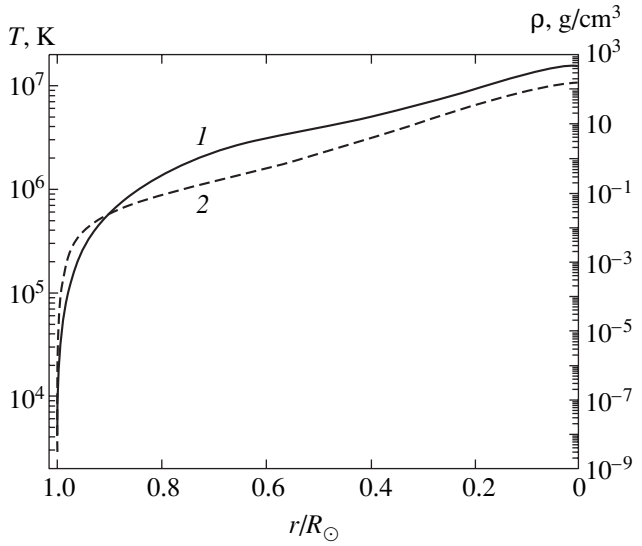


Fig. 3. Solar trajectory: (*1*) temperature (K) and (*2*) density (g/cm^3) as functions of the relative radius $r/R_\odot$.

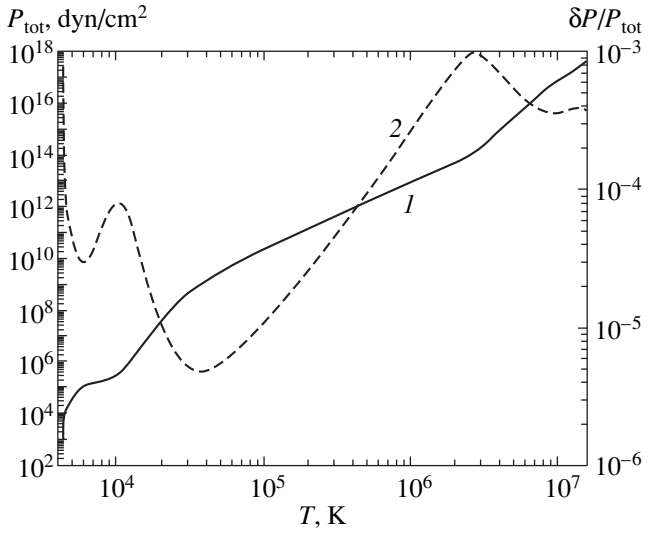


Fig. 4. (*1*) Total pressure and (*2*) fraction of radiation pressure in the total pressure as functions of temperature along the solar trajectory.

and (185), which are expansions in powers of the Born parameter [$\sim e^2/(\hbar v_T)$],

$$\varkappa_{ep}^3\Sigma_{SRM}^{BS} = \frac{\pi^4}{3}\left(\frac{e^2}{T}\right)^4\frac{1}{\varkappa_{ep}} - \frac{64\pi^2}{15}\zeta(3)\left(\frac{e^2}{T}\right)^5\frac{1}{\varkappa_{ep}^2} + \frac{8\pi^3}{3}\zeta(4)\left(\frac{e^2}{T}\right)^6\frac{1}{\varkappa_{ep}^3} - \ldots, \quad (189)$$

$$\varkappa_{ep}^3\Sigma_{\text{P–L}}^{BS} = \frac{\pi^4}{12}\left(\frac{e^2}{T}\right)^4\frac{1}{\varkappa_{ep}} - \frac{\pi^3}{6}\zeta(4)\left(\frac{e^2}{T}\right)^6\frac{1}{\varkappa_{ep}^3} + \ldots, \quad (190)$$

$$\varkappa_{ep}^3\Sigma^{SS} = -\frac{7\pi^4}{24}\left(\frac{e^2}{T}\right)^4\frac{1}{\varkappa_{ep}} + \frac{68\pi^2}{15}\zeta(3)\left(\frac{e^2}{T}\right)^5\frac{1}{\varkappa_{ep}^2} - \frac{31\pi^3}{12}\zeta(4)\left(\frac{e^2}{T}\right)^6\frac{1}{\varkappa_{ep}^3} - \ldots, \quad (191)$$

$$\varkappa_{ep}^3(\Sigma^{SS})' = -\frac{\pi^4}{24}\left(\frac{e^2}{T}\right)^4\frac{1}{\varkappa_{ep}} + \frac{4\pi^2}{15}\zeta(3)\left(\frac{e^2}{T}\right)^5\frac{1}{\varkappa_{ep}^2} - \frac{\pi^3}{12}\zeta(4)\left(\frac{e^2}{T}\right)^6\frac{1}{\varkappa_{ep}^3} - \ldots. \quad (192)$$

A comparison of (189) and (191) shows that the principal nonvanishing contribution of bound states is commensurate in magnitude to the contribution of scattering states, whereas the first term in (190) is two times larger in magnitude than the corresponding term in (192). Note that Eq. (192) was obtained in [10].

The chemical picture that includes the ideal gas, Debye–Hückel, and bound state contributions to the equation of state is often used in the literature. All corrections, including those for diffraction, screening by "dressed" particles, and interactions in the continuum (arising from scattering states containing "classical" logarithmic contributions and quantum sums in the Born parameter $e^2/\hbar v_T$), are then ignored. This is justified in the limit $I/T \gg 1$, because the main contribution to the equation of state is then made by bound states and is proportional to $\exp(I/T)$. In the center of the Sun, we have $I/T \ll 1$. For this reason, the neglect of scattering states while the contribution of bound states is retained is illegal. However, as both contributions (Σ values) are then small ($\sim 10^{-4}$) and the contribution of bound states predominates at the periphery, the chemical picture may be fairly "correct."

The complete physical model for determining the equation of state of a hydrogen plasma along the solar trajectory is compared with the chemical models represented by Σ_{SRM}^{BS} or $\Sigma_{\text{P–L}}^{BS}$ in Section 7.

The equation of state of a plasma that describes pressure $P(\rho, T)$ and other thermodynamic functions should be augmented by the contribution of equilibrium thermal radiation in the plasma. Equations for radiation intensity in an absorbing medium were obtained in [42]. It follows from these equations that, with respect to the $\varepsilon''/\varepsilon' \ll 1$ parameter, we can use equations

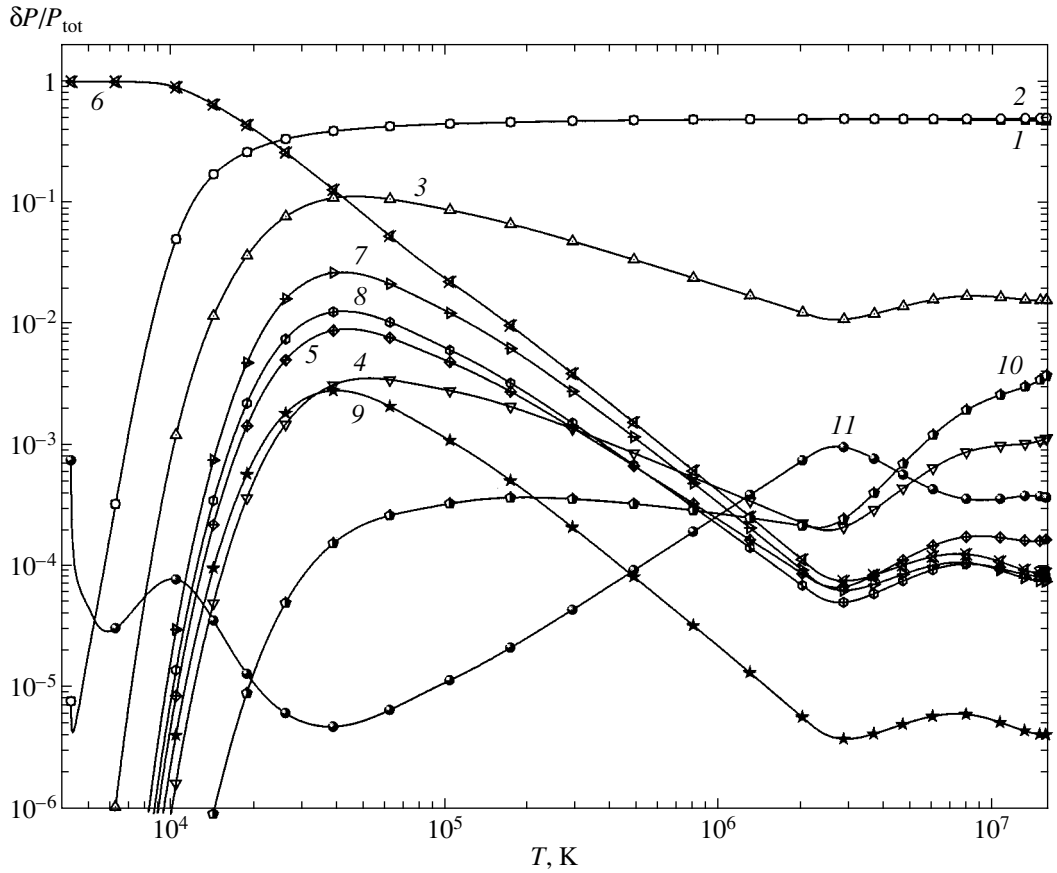


Fig. 5. Absolute values of relative contributions $\delta P/P_{\rm tot}$ to the total pressure versus temperature along the solar trajectory ($\delta P = -\delta\Omega/V$). Components: (*1*) $P_{0,\,i}$, (*2*) $P_{0,\,er}$, (*3*) $\delta P_{\rm D-H}$, (*4*) $\Delta\Omega_{\rm diff}/V$, (*5*) $\delta\Omega^{cl}/V$, (*6*) $-\delta\Omega^{BS}_{SRM}/V$, (*7*) $\delta\Omega^{q}_{ep}/V$, (*8*) $-\delta\Omega^{q}_{pp}/V$, (*9*) $-\delta\Omega^{q}_{ee}/V$, (*10*) $-\delta\Omega^{\rm exch}_{ee}/V$, and (*11*) P_R.

for radiation in a transparent medium. For instance, we have $\bar{k}_\omega c/\bar{\omega} \sim 10^{-6}$ in the center of the Sun according to the *S*-model [1]; here, $\bar{k}_\omega$ is the mean absorption coefficient (opacity), c is the velocity of light, and $\bar{\omega}$ is the mean frequency of thermal radiation.

Let us write an equation for energy in a transparent dense plasma [43] taking into account the relation $k = n\omega/c$ ($n = \sqrt{1-\omega_p^2/\omega^2}$ is the refractive index of the plasma),

$$E_R = 2V\int \frac{\hbar\omega}{\exp(\hbar\omega/T)-1}\frac{d\mathbf{k}}{(2\pi)^3} = \frac{\hbar V}{\pi^2 c^3}\int\limits_{\omega_p}^{\infty}\frac{(\omega^2-\omega_p^2)^{1/2}\omega^2}{\exp(\hbar\omega/T)-1}d\omega. \quad (193)$$

Here,

$$\omega_p^2 = 4\pi e^2 n_e/m_e$$

is the electronic plasma frequency and V is the volume of the system ($V \longrightarrow \infty$). It is taken into account in (193) that radiation with frequencies $\omega < \omega_p$ does not propagate in the system as free radiation. The expansion of the denominator in (193) in powers of the exponential function $\exp(-\hbar\omega/T)$ yields

$$\frac{E_R}{V} = \frac{\hbar\omega_p T^3}{\pi^2 c^3 \hbar^3} \times \sum_{n=1}^{\infty}\frac{1}{n^3}[zK_0(z) + 2zK_2(z) + (2+z^2)K_1(z)], \quad (194)$$

where $K_n(z)$ stands for the Macdonald functions [38] and $z = n\hbar\omega_p/T$. In the limit $\hbar\omega_p \ll T$ ($\hbar\omega_p/T \sim 1/4$ in the center of the Sun), (194) yields (cf. [28])

$$\frac{E_R}{V} = \frac{\pi^2 T^4}{15\hbar^3 c^3}\left[1 - \frac{5}{4\pi^2}\left(\frac{\hbar\omega_p}{T}\right)^2\right]. \quad (195)$$

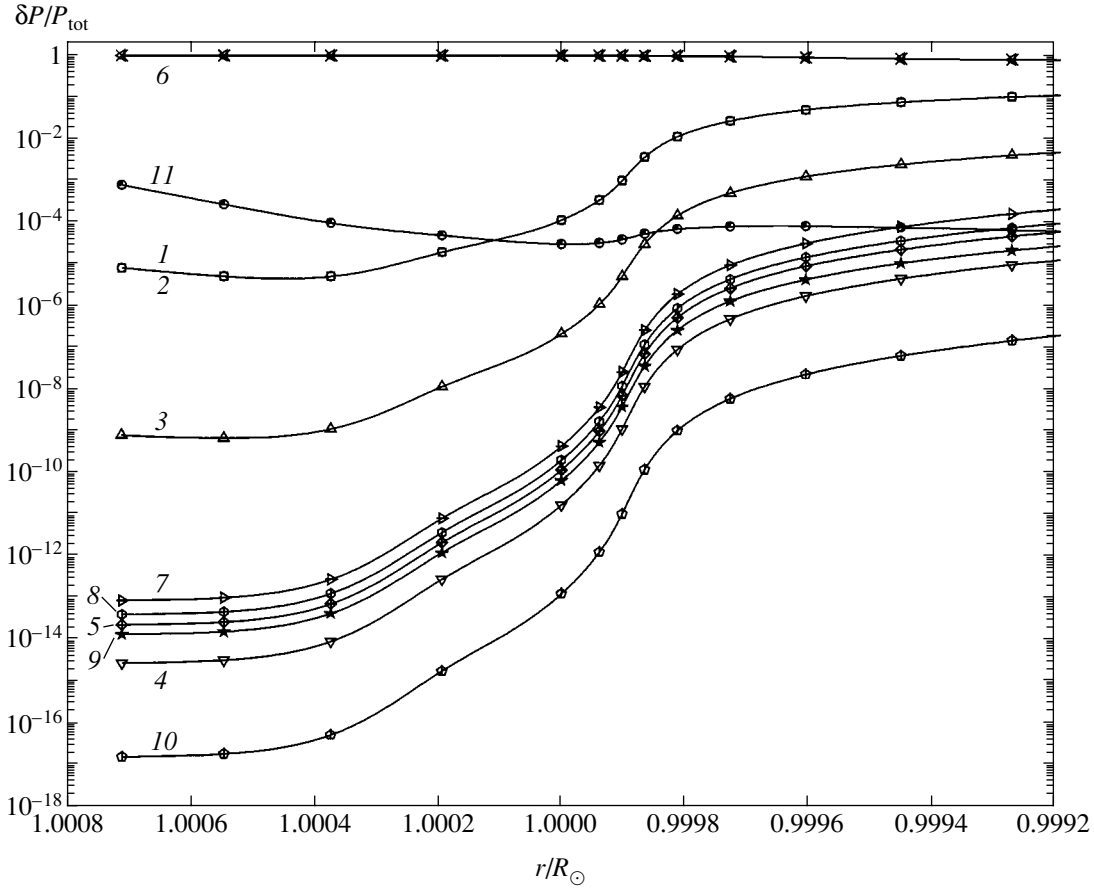


Fig. 6. Absolute values of relative contributions $\delta P/P_{\text{tot}}$ to the total pressure versus relative radius $r/R_\odot$ at the periphery of the Sun ($\delta P = -\delta\Omega/V$). Components: (*1*) $P_{0,\,i}$, (*2*) $P_{0,\,er}$, (*3*) $\delta P_{\text{D–H}}$, (*4*) $\Delta\Omega_{\text{diff}}/V$, (*5*) $\delta\Omega^{cl}/V$, (*6*) $-\delta\Omega^{BS}_{SRM}/V$, (*7*) $\delta\Omega^{q}_{ep}/V$, (*8*) $-\delta\Omega^{q}_{pp}/V$, (*9*) $-\delta\Omega^{q}_{ee}/V$, (*10*) $-\delta\Omega^{\text{exch}}_{ee}/V$, and (*11*) P_R.

In a similar way, we obtain free radiation energy in the form

$$F_R = \frac{TV}{\pi^2 c^3}\int_{\omega_p}^{\infty}\omega\sqrt{\omega^2-\omega_p^2}\ln\left(1-\exp\left(-\frac{\hbar\omega}{T}\right)\right)d\omega$$
$$= -\frac{\hbar V}{3\pi^2 c^3}\int_{\omega_p}^{\infty}\frac{(\omega^2-\omega_p^2)^{3/2}}{\exp(\hbar\omega/T)-1}d\omega \qquad (196)$$
$$= -\frac{\hbar V\omega_p^2 T^2}{\pi^2 c^3\hbar^2}\sum_{n=1}^{\infty}\frac{1}{n^2}K_2(z).$$

As

$$\frac{\partial\omega_p}{\partial V} = -\frac{\omega_p}{2V}$$

we can use (196) to determine radiation pressure,

$$P_R = -\left(\frac{\partial F_R}{\partial V}\right)_T = \frac{\omega_p^2 T^2}{\hbar\pi^2 c^3}\sum_{n=1}^{\infty}\frac{1}{n^2}K_2(z)$$
$$+\frac{\omega_p^3 T}{2\pi^2 c^3}\sum_{n=1}^{\infty}\frac{1}{n}K_1(z). \qquad (197)$$

In the limit $\hbar\omega_p \ll T$, (197) yields [28]

$$P_R^0 = \frac{T^4\pi^2}{45\hbar^3 c^3}. \qquad (198)$$

To determine the velocity of sound along the solar trajectory, we must know the c_V^R value (the heat capacity of radiation),

$$c_V^R = \frac{\partial E_R/V}{\partial T},$$

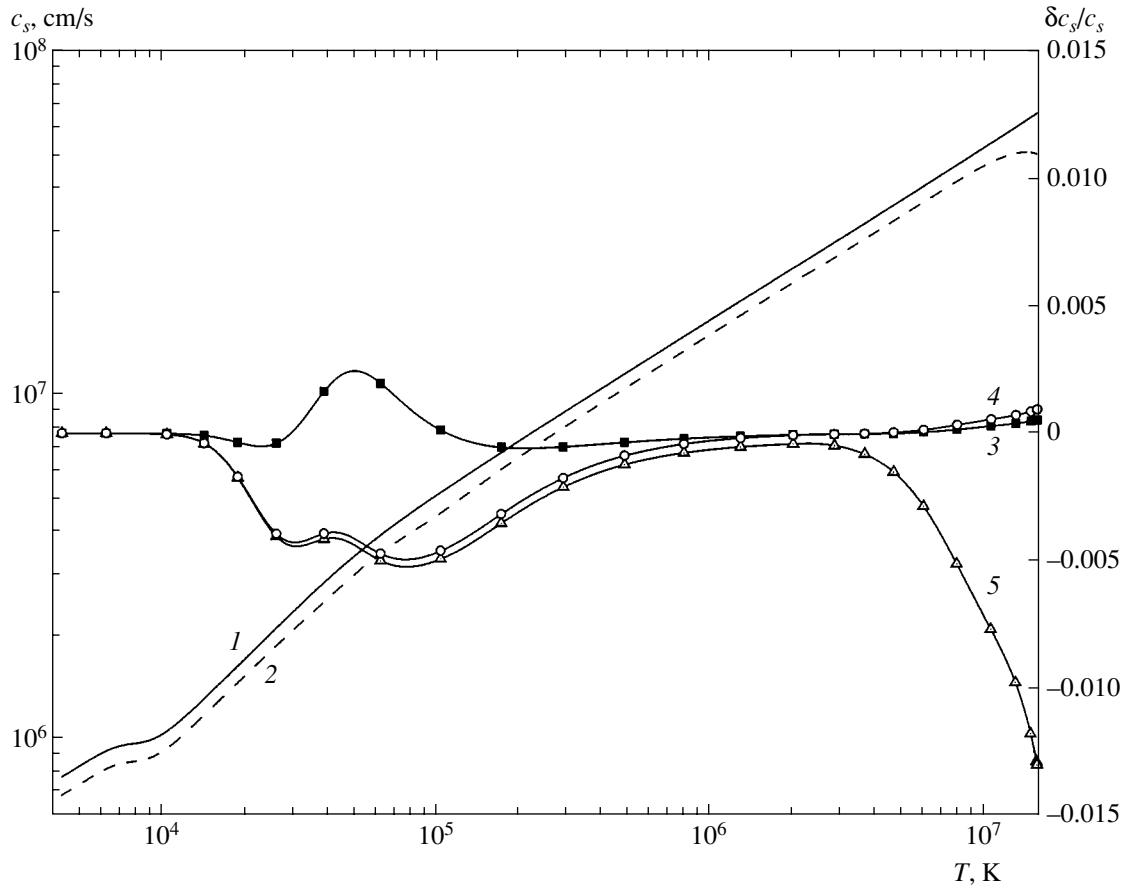


Fig. 7. Velocity of sound ((*1*) hydrogen physical model and (*2*) *S*-model [1]) and relative deviations from (*1*) for various chemical models taking into account radiation [(*3*) $\Sigma^{BS}_{\text{P–L}}$ and (*4*) Σ^{BS}_{SRM} models and (*5*) Σ^{BS}_{SRM} model ignoring degeneracy] as functions of temperature along the solar trajectory.

and the derivatives $(\partial P_R/\partial T)_\rho$ and $(\partial P_R/\partial \rho)_T$. The corresponding equations are easily obtained from (194) and (197). We omit the resulting fairly cumbersome formulas.

Note that equilibrium between radiation and matter presupposes that the condition $\bar{k}_\omega R \gg 1$ is fulfilled. At the periphery, in the photosphere, this condition is violated [1]. For instance,

$$\rho \sim 4 \times 10^{-9} \text{ g/cm}^3,$$

$$\bar{k} = \kappa\rho = 8 \times 10^{-3} \times 4 \times 10^{-9} = 3.2 \times 10^{-11} \text{ cm}^{-1},$$

$$R_\odot \approx 7 \times 10^{10} \text{ cm}, \quad \bar{k}R_\odot \sim 2.$$

Solving direct helioseismology problems requires the determination of various thermodynamic values along the solar trajectory. We restrict our consideration to a hydrogen plasma (whereas the Sun contains other elements too, largely He and also Li, Be, B, C, N, O, Ne, …, Fe) because the generalization of equations of type (104) with Σ^{BS} in one or another form (even within the framework of the chemical model) to multielectron bound states requires solving the problem of three and more bodies. Strict results have not been obtained thus far even for HeI, and the problem requires further inquiries. Several comments will be made in Conclusions.

The velocity of sound in a hydrogen plasma can be calculated by the equation [44]

$$c_S^2 = \left(\frac{\partial P}{\partial \rho}\right)_T \left[1 + \frac{T\left(\frac{\partial P}{\partial T}\right)_\rho^2}{c_V \rho \left(\frac{\partial P}{\partial \rho}\right)_T}\right]. \tag{199}$$

Here, pressure P and heat capacity c_V are represented by the sums of the contributions of matter P_M (c_V^M) and radiation P_R (c_V^R) and ρ is the density of the plasma.

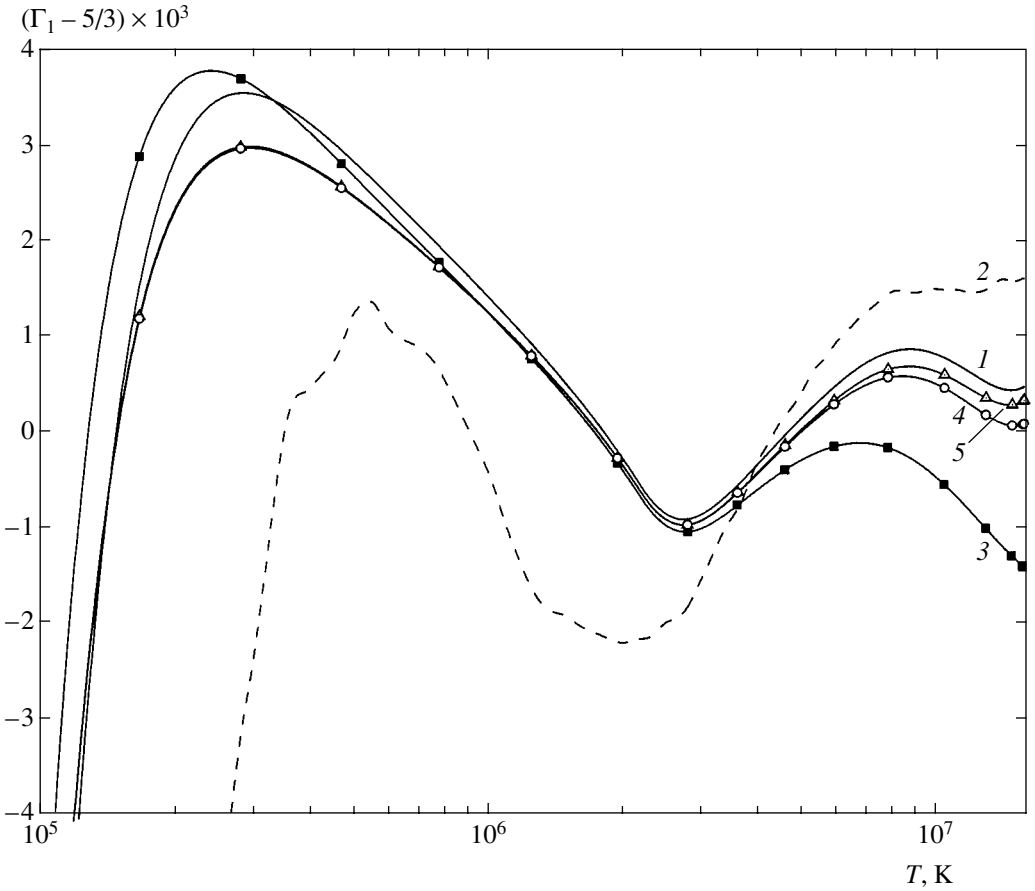


Fig. 8. Relative deviations of the adiabatic compressibility index from the ideal gas value $[(\Gamma_1 - 5/3) \times 10^3]$ depending on temperature along the solar trajectory: (*1*) hydrogen physical model, (*2*) *S*-model [1], and chemical models taking radiation into account [(*3*) $\Sigma^{BS}_{\text{P–L}}$, (*4*) Σ^{BS}_{SRM}, and (*5*) Σ^{BS}_{SRM} with ignoring degeneracy].

Another important parameter used in helioseismology is the adiabatic exponent Γ_1,

$$\Gamma_1 = \frac{\rho}{P} c_S^2. \tag{200}$$

7. NUMERICAL CALCULATIONS OF THE EQUATION OF STATE OF A WEAKLY NONIDEAL HYDROGEN PLASMA ALONG THE SOLAR TRAJECTORY

Equations (5), (6), (9), (179), (180), (182), (16), (52), (57), (98), (104), (156), (163), (160), (164), (165), (170), (171), (194), and (197) were used to calculate the equation of state of a weakly nonideal hydrogen plasma, that is, the dependence of the total pressure $P(r)$ or $P(T)$ along the solar trajectory within the framework of the *S*-model [1].

Density and temperature distributions as functions of the current radius counted from the center of the Sun [1] are shown in Fig. 3. Calculations taking into account all the contributions specified above characterize the physical picture, which, according to thermodynamic perturbation theory, is represented with accuracy to the second virial coefficient (up to ζ^2) for the hydrogen plasma.

The total pressure of the plasma and radiation along the solar trajectory is shown in Fig. 4, where we also see the fraction of radiation pressure with respect to the total pressure (approximately 4×10^{-4} in the center of the Sun).

The absolute values of relative contributions $\delta P/P_{\text{tot}}$, where $\delta P = -\delta\Omega/V$, to the total pressure are shown in Figs. 5 (versus temperature) and 6 (at the periphery, versus relative radii $1.0008 \geq r/R_\odot \geq 0.9992$). We see that the contribution of the pressure of the electrons and protons, which approximately equals 1 in the temperature coordinates, sharply drops at the periphery approximately to 10^{-5}. At the periphery, bound states (atoms) predominate. Their contribution in the center of the Sun, like that of *e–p* scattering states, is approximately 10^{-4}. The contribution of Coulomb interaction (on the order of 10^{-2} in the center of the Sun) reaches a maximum (about 10^{-1}) at $T \sim 5 \times 10^4$ K and decreases to 10^{-9} at the periphery. Diffraction corrections (on the order of

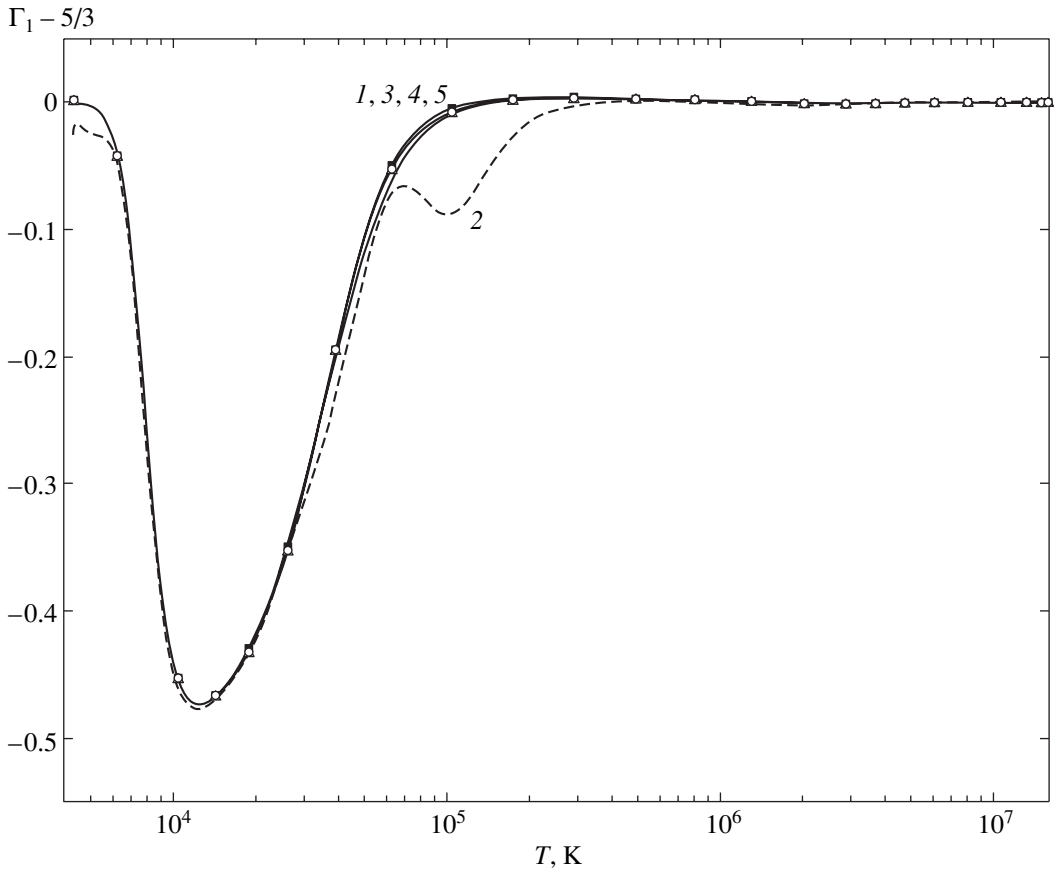


Fig. 9. Relative deviations of the adiabatic compressibility index from the ideal gas value ($\Gamma_1 - 5/3$); see Fig. 8 for notation.

10^{-3} in the center of the Sun) pass a maximum (about 4×10^{-3}) at $T \approx 5 \times 10^4$ K and then, like all interactions involving charged particles (the contributions of the summed logarithm, scattering states arising from *e–p*, *p–p*, and *e–e* interactions, and *e–e* exchange interaction) sharply drop at the periphery.

The relative contribution of the total logarithmic term is about 2×10^{-4} in the center of the Sun (accordingly, the correction for screening by "dressed" particles is an order of magnitude smaller).

Exchange interaction in the center of the Sun, next in significance to Coulomb contribution, is approximately 4×10^{-3}. The radiation pressure is about 4×10^{-4}, and the contribution of *p–p* scattering states is on the order of 10^{-4} in the center of the Sun. The smallest contribution of *e–e* scattering states is approximately 4×10^{-6}.

We also see that many of the corrections to the second virial coefficient are quite substantial, taking into account the high accuracy of the inversion procedure in helioseismology problems.

We must again mention that we only consider the contribution of a hydrogen plasma on the assumption (obviously incorrect, especially close to the center of the Sun, because of the presence of He and other elements) that this is hydrogen that is responsible for the total density of matter. In spite of the exceedingly conventional character of this model, the purpose of our calculations was to qualitatively estimate the relative contributions of various total plasma pressure components. At the same time, the accuracy of this model in the usual sense (leaving precision helioseismology aside) is quite reasonable in comparison with the data of the complete *S*-model, including He and other elements distributed along the solar trajectory. For instance, the distributions of the velocity of sound c_S on the temperature scale calculated within the framework of the hydrogen physical model [see (199)] and the S-model are shown in Fig. 7. The same figure contains the relative deviations of the velocity of sound calculated using various chemical models. The chemical models ignore all corrections except the contribution of free particles, the Debye–Hückel Coulomb interaction, the contribution of bound states according to the Planck–Larkin equation ($\Sigma^{BS}_{\text{P–L}}$) or represented by Σ^{BS}_{SRM}, and the contribution of radiation. The difference between the results obtained using two methods for the inclusion of bound states is approximately 7×10^{-3} at $T \sim 10^5$ K, when degeneracy effects are inessential. In

the center of the Sun, the role played by degeneracy is noticeable, and the two models, $\Sigma^{BS}_{\text{P–L}}$ and Σ^{BS}_{SRM}, one taking degeneracy into account and the other ignoring it, are very close to each other, but the relative deviations of the models including degeneracy from those that ignore it are larger than 10^{-2}.

The deviations of the adiabatic exponent from its ideal gas value $(\Gamma_1 - 5/3) \times 10^3$ [see (200)] along the solar trajectory are shown in Fig. 8. We can compare the data calculated according to the complete physical model and chemical models based on equations for $\Sigma^{BS}_{\text{P–L}}$ and Σ^{BS}_{SRM} (with and without taking degeneracy into account). The same figure contains the results of calculations by the *S*-model for the "real" Sun containing all elements [1]. Even though the hydrogen model is obviously incomplete, the difference between the $(\Gamma_1 - 5/3) \times 10^3$ value calculated with its use and the value obtained according to the *S*-model is not very large, especially close to the center of the Sun, where the physical model and the Σ^{BS}_{SRM} variant of the chemical model are closer to experiment than the $\Sigma^{BS}_{\text{P–L}}$ chemical model (which even gives deviations opposite in sign).

The $\Gamma_1 - 5/3$ value is plotted in the entire range along the coordinate axes in Fig. 9, which gives an additional idea of the quality of the hydrogen model. We see that the adiabatic exponent decreases in the region of hydrogen ionization. The second, smaller, peak present in *S*-model calculations is caused by the ionization of helium, which is absent in our model. The difference between the chemical models specified above is unnoticeable on the scale of the figure.

8. CONCLUSIONS

The development of helioseismology requires knowledge of the equation of state with high accuracy. We therefore performed an independent and detailed derivation of a converging equation of state taking into account relativistic ($\sim T/(mc^2)$) and degeneracy ($\sim n_e \lambda\!\!\!^{-3}_e$) effects and consistently including the neutrality condition for activities $\zeta_e = \zeta_p$, Coulomb interaction in the Debye–Hückel approximation ($\sim\Gamma_{\text{D}}$) with various corrections on the order of $\Gamma_{\text{D}}\lambda\!\!\!^{-}\kappa$, and contributions to the second virial coefficient from the converging logarithmic term $[\Gamma^2_{\text{D}} \ln(m_p/m_e)]$, bound states ($\sim\zeta_e \lambda\!\!\!^{-3}_{ep} \Sigma^{BS}_{SRM}$), scattering states arising from interactions of all types ($\sim\zeta_e \lambda\!\!\!^{-3}_{ij} \Sigma^{SS}_{ij}$), exchange interaction ($\sim\zeta_e \lambda\!\!\!^{-2}_e e^2/T$), and radiation pressure taking into account plasma effects on radiation dispersion properties. Numerical calculations of the equation of state, the velocity of sound, and the adiabatic exponent were performed for a hydrogen plasma along the solar trajectory with the use of the complete physical model and various chemical models. Note that, at the level of the current accuracy of measurements in helioseismology, qualitative sorting out of certain approximations is possible in principle (for instance, the results obtained in [6] are closer to experiment than those reported in [4]).

Note several problems that require solution to promote further developments in this field of study. First and foremost, we must consistently go beyond the scope of the approximation that describes the equation of state up to the second virial coefficient. Even taking into account the contributions of the H^-, H^+_2, and H_2 hydrogen states to pressure falls outside the approximation limited by the second virial coefficient. We already mentioned the problem of describing the contribution of many-electron states of neutral and charged species (HeI, HeII, etc. over all the elements) and many-body scattering states. Heroic efforts made in this direction [45, 46] cannot be considered conclusive. Indeed, the method [45] for taking into account connected diagrams (P_{CD}) when simple irreducible diagrams

$$P_{SD} = TS(\zeta_k)$$

(also see [13]) are known, which is based on the equation

$$P_{CD} = T\sum_k \frac{\zeta_k}{2}\left(\frac{\partial S}{\partial \zeta_k}\right)^2, \qquad (201)$$

itself leads to inaccuracies because it involves the integration in charge in quantum thermodynamic perturbation theory [24]. For instance, with $S(\zeta_k)$ in the Debye–Hückel approximation,

$$S = \frac{\kappa^3_{\text{D}}}{12\pi}, \qquad (202)$$

we obtain from (201)

$$P_{CD} = \frac{\pi}{2} T\left(\frac{e^2}{T}\right)^3 \left(\sum_i \zeta_i z_i^4\right)\left(\sum_j \zeta_j z_j^2\right), \qquad (203)$$

which differs from the "correct" result [Eq. (57)] in the coefficient. Because of the integration in charge, we here have $\pi/2$ instead of $\pi/3$ ([8] contains a similar slip).

At the same time, the approach suggested in [45, 46] allows us to at least qualitatively predict the structure of the converging partition function of a multielectron atom. Note that equations of the type of the Planck–Larkin equation or the equation for Σ_{SRM} diverge if literally generalized, for instance, to a two-electron (helium-like) atom (ion) because of the presence of doubly excited states. Let us consider the limit of a large nuclear charge ($z \gg 1$), when electron–electron interaction can be ignored in the principal approxima-

tion compared with electron–nucleus interaction. Taking into account that [cf. (104)]

$$S_{SD} = \zeta_e \zeta_p T \chi_{ep}^3 \Sigma^{BS}$$

we obtain from (201)

$$P_{CD} = \frac{T}{2} \zeta_p \zeta_e^2 \chi_{ep}^6 (\Sigma_{ep}^{BS})^2. \quad (204)$$

This result appears to be reasonable, at least qualitatively. To substantiate equations of this type, we must use the Faddeev equations for the problem of three bodies [47].

Also note that the Sun is a unique scientific laboratory. Indeed, first-principle calculations of many processes in it are possible because its plasma is weakly nonideal. For instance, in the center of the Sun, the condition

$$T \gg \max(e^2 \kappa_D, \hbar \omega_p, \mathrm{Ry}, E_{St} d) \quad (205)$$

is satisfied (here, $E_{St} = e n_e^{2/3}$ is the Stark microfield in the plasma and $d \sim e a_0$). For this reason, the principal cut-off factor for the partition function of an atom is temperature, because the contribution of bound states is on the order of $(\mathrm{Ry}/T)^2$ [see (100), (101)]. In (205), $E_{St} d$ is the interaction energy between the Stark microfield and the atomic dipole $d \sim e a_0$. The presence of the large parameter given by (205) allows us to calculate the equation of state starting with the ideal gas approximation and sequentially introduce the other effects, including the influence of the discrete spectrum on the pressure of a plasma, where unperturbed atomic states arise at the first step. A qualitative consideration based on the decay of "atoms" caused by screening, broadening, ionization under the action of microfields (the Unsöld mechanism [4]), etc. leads to a picture different from that obtained using perturbation theory with the expansion in powers of activities, which is even not substantiated by experiments. (Cf. [1], where it was noted that the results of [6] are closer to the inversion data than those obtained in [4].) Such a literal taking into account of Debye screening in calculations of atomic states (e.g., see [13, 14]) causes large shifts of atomic levels and, accordingly, the centers of spectral lines, which are not observed in weakly nonideal plasmas of arc discharges [48, 49].

In our view, taking into account broadening effects (see (116) and the results reported in [20]) in describing both spectral lines and the thermodynamic properties of a weakly nonideal dense plasma is a more promising approach. In principle, it binds up the problems of radiation gas dynamics with those of radiation-collisional kinetics, in which "atoms" are represented differently in pressure and radiation calculations [20].

ACKNOWLEDGMENTS

The authors thank S.V. Ayukov, V.A. Baturin, V.K. Gryaznov, I.L. Iosilevskiĭ, T. Kato, R. More, and V.E. Fortov for many fruitful discussions. This work was financially supported by the Russian Foundation for Basic Research (project no. 04-02-16775-a) and a grant from the President of the Russian Federation for support of leading scientific schools (no. NSh 1257.2003.2).

REFERENCES

1. J. Christensen-Dalsgaard, W. Däppen, W. A. Dziembowski, and J. A. Guzik, in *Proceedings of NATO ASI on Variable Stars as Important Astrophysical Tools*, Ed. by C. Ibanoglu (Kluwer, Dordrecht, 2000), p. 59; http://bigcat.ifa.au.dk/~jcd/solar_models/cptrho.15bi.d.15c.
2. V. A. Baturin, W. Däppen, D. O. Gough, and S. V. Vorontsov, Mon. Not. R. Astron. Soc. **316**, 71 (2000).
3. S. Basu, W. Däppen, and A. Nayfonov, Astrophys. J. **518**, 985 (1999).
4. D. G. Hummer and D. Mihalas, Astrophys. J. **331**, 794 (1988).
5. D. Mihalas, W. Däppen, and D. G. Hummer, Astrophys. J. **331**, 815 (1988).
6. F. J. Rogers, F. J. Swenson, and C. A. Iglesias, Astrophys. J. **456**, 902 (1996).
7. A. A. Vedenov and A. I. Larkin, Zh. Éksp. Teor. Fiz. **36**, 1133 (1959) [Sov. Phys. JETP **9**, 806 (1959)].
8. A. I. Larkin, Zh. Éksp. Teor. Fiz. **38**, 1896 (1960) [Sov. Phys. JETP **11**, 1363 (1960)].
9. Yu. G. Krasnikov, Zh. Éksp. Teor. Fiz. **53**, 2223 (1967) [Sov. Phys. JETP **26**, 1246 (1968)].
10. V. P. Kopyshev, Zh. Éksp. Teor. Fiz. **55**, 1304 (1968) [Sov. Phys. JETP **28**, 684 (1969)].
11. W. Ebeling, H.-J. Hoffman, and G. Kelbg, Beitr. Plasmaphys. **7**, 233 (1967).
12. W. Ebeling, G. Kelbg, and K. Rohde, Ann. Phys. (Leipzig) **21**, 235 (1968).
13. W. Ebeling, W.-D. Kraeft, and D. Kremp, *Theory of Bound States and Ionization Equilibrium in Plasmas and Solids* (Akademie, Berlin, 1976; Mir, Moscow, 1979).
14. W.-D. Kraeft, D. Kremp, W. Ebeling, and G. Röpke, *Quantum Statistics of Charged Particle Systems* (Plenum, New York, 1986; Mir, Moscow, 1988).
15. C. A. Rouse, Astrophys. J. **272**, 377 (1983).
16. C. E. Moore, M. G. J. Minnaert, and J. Houtgast, *The Solar Spectrum 2935 Å to 8770 Å* (1966), NBS Monograph, Vol. 61.
17. A. Unsöld, Z. Astrophys. **24**, 355 (1948).
18. G. É. Norman and A. N. Starostin, Teplofiz. Vys. Temp. **8**, 413 (1970) [Sov. Phys. High Temp. **8**, 381 (1970)].
19. V. S. Vorob'ev and A. L. Khomkin, Teor. Mat. Fiz. **8**, 109 (1971).
20. A. N. Starostin, V. C. Roerich, and R. M. More, Contrib. Plasma Phys. **43** (5–6), 369 (2003).

21. E. Uhlenbeck and G. E. Beth, Physica (The Hague) **3**, 729 (1936); G. E. Beth and E. Uhlenbeck, Physica (The Hague) **4**, 915 (1937).
22. N. Levinson, K. Dan. Vidensk. Selsk. Mat. Fys. Medd. **25** (9), 1 (1949).
23. T. Matsubara, Prog. Theor. Phys. **14**, 351 (1955).
24. 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).
25. L. Kadanoff and G. Baym, *Quantum Statistical Mechanics* (Benjamin, New York, 1962).
26. L. V. Keldysh, Zh. Éksp. Teor. Fiz. **47**, 1515 (1964) [Sov. Phys. JETP **20**, 1018 (1965)].
27. E. M. Lifshitz and L. P. Pitaevskiĭ, *Physical Kinetics* (Nauka, Moscow, 1979; Pergamon Press, Oxford, 1981).
28. 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.
29. V. M. Galitskiĭ, Zh. Éksp. Teor. Fiz. **34**, 151 (1958) [Sov. Phys. JETP **7**, 104 (1958)].
30. E. W. Montroll and J. C. Ward, Phys. Fluids **1**, 55 (1958).
31. H. E. DeWitt, J. Math. Phys. **3**, 1216 (1962); **7**, 616 (1966).
32. E. S. Fradkin, in *Quantum Field Theory and Hydrodynamics* (Nauka, Moscow, 1965), Tr. Fiz. Inst. im. P. N. Lebedeva, Akad. Nauk SSSR, Vol. 29 [in Russian].
33. A. A. Likalter, Zh. Éksp. Teor. Fiz. **56**, 240 (1969) [Sov. Phys. JETP **29**, 133 (1969)].
34. V. K. Gryaznov, I. L. Iosilevskiĭ, and V. E. Fortov, Prikl. Mekh. Tekh. Fiz. **3**, 70 (1973).
35. A. N. Starostin, in *Proceedings of IX International Conference on Phenomena in Ionized Gases* (Bucharest, 1969), p. 366.
36. 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).
37. V. A. Fock, Izv. Akad. Nauk SSSR, Otd. Mat. Estest. Nauk **2**, 169 (1935); Z. Phys. **98** (3/4), 145 (1935).
38. I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products*, 5th ed. (Nauka, Moscow, 1971; Academic, New York, 1980).
39. A. I. Akhiezer and V. B. Berestetskiĭ, *Quantum Electrodynamics*, 3rd ed. (Nauka, Moscow, 1969; Wiley, New York, 1965).
40. *Higher Transcendental Functions (Bateman Manuscript Project)*, Ed. by A. Erdélyi (McGraw-Hill, New York, 1953; Nauka, Moscow, 1965), Vol. 1.
41. A. P. Prudnikov, Yu. A. Brychkov, and O. I. Marichev, *Integrals and Series, Elementary Functions* (Nauka, Moscow, 1981; Gordon and Breach, New York, 1986).
42. Yu. K. Zemtsov, A. Yu. Sechin, and A. N. Starostin, Zh. Éksp. Teor. Fiz. **110**, 1654 (1996) [JETP **83**, 909 (1996)].
43. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media*, 2nd ed. (Nauka, Moscow, 1982; Pergamon, Oxford, 1984).
44. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics*, 3rd ed. (Nauka, Moscow, 1986; Pergamon, New York, 1987).
45. F. J. Rogers and H. DeWitt, Phys. Rev. A **8**, 1061 (1973).
46. F. J. Rogers, Phys. Rev. A **10**, 2441 (1974).
47. L. D. Faddeev, Zh. Éksp. Teor. Fiz. **39**, 1459 (1960) [Sov. Phys. JETP **12**, 1014 (1961)].
48. H. R. Griem, Phys. Rev. **131**, 1170 (1963).
49. W. L. Wiese, D. E. Kelleher, and D. R. Paquette, Phys. Rev. A **6**, 1132 (1972).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 100, No. 1, 2005, pp. 199–207.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 1, 2005, pp. 220–229.

STATISTICAL, NONLINEAR, AND SOFT MATTER PHYSICS

Bistable Switching in Chiral Nematic Liquid Crystal Layers with a 2π-Twist Ground State

S. P. Palto and M. I. Barnik

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

Received July 15, 2004

Abstract—We present experimental evidence and theoretical justification for the phenomenon of bistable electrooptical switching in nematic liquid crystals doped with chiral compounds so that the ground state corresponds to a 360° helical twist of the director. The system exhibits switching between the ground and a metastable unwound states. The observed effect differs from the well-known Berreman–Heffner bistable switching by the absence of a topologically nonequivalent ground state of the liquid crystal layer with a 180° twist of the director. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The physics of formation of stable and metastable states in liquid crystals (LCs) and the effects of switching between these states constantly attract the attention of researchers. This research activity is related both to the interest in elucidating special features in the interaction of LC molecules with the alignment surfaces and to the new prospects of using the bistable switching effects in electrooptical devices and display technologies [1–4].

One of the first effects of bistable switching in nematic liquid crystals (NLCs) was demonstrated in the early 1980s by Berreman and Heffner [5, 6]. It was shown that NLCs doped to a certain extent with a chiral additive are capable of featuring transitions to metastable states under the action of an applied electric field. Because of some features of the interaction between the LC director and hydrodynamic back-flows arising in the course of LC reorientation in the applied field, it was also possible to provide for reversible switching between these optically distinguishable states with different director twist angles. A distinctive feature of the Berreman–Heffner effect is that, for the typical boundary conditions with a small director pretilt angle at the NLC layer boundaries, the ground state is characterized by a helix with a 180° twist of the director, while the switching takes place between the 0° and 360° twist states. Thus, the system in fact features three states, which leads to problems in practical applications of this phenomenon.

The Berreman–Heffner effect had been extensively studied until quite recently [7–10]. These investigations were aimed, in particular, at finding conditions for effectively eliminating the ground 180° twist state. This is related to the fact that the 180° state is topologically nonequivalent to the switched (0° and 360°) states. If the driving field is absent for a sufficiently long time, the director exhibits relaxation to the ground state via defects. Elimination of the defects requires using special field configurations and/or special boundary conditions [7, 10]. For symmetric boundary conditions, nondefect transitions are possible only between topologically equivalent states even in the material with a weak anchoring energy [11]. A special case is realized under asymmetric boundary conditions with a weak anchoring at one of the boundaries, whereby transitions are possible between the 0° and 180° twist states [4]. However, the creation of controlled asymmetric anchoring conditions is a rather difficult technological problem.

This paper demonstrates the possibility of bistable switching under the conditions where the minimum free energy of an NLC layer corresponds to a helical state of the director with a 360° twist. In the case under consideration, this is the ground state and it is topologically equivalent to the second switched (unwound) state, so that the problem of the third (topologically nonequivalent) state is eliminated.

The paper consists of theoretical and experimental parts. The theoretical part reports on the results of numerical simulation of the LC director dynamics and describes the requirements to parameters that the LC materials have to obey in order to provide bistable switching between the 360° twist ground state and a metastable homogeneous unwound state. The experimental part presents evidence of the actual bistable switching in NLCs of two types. The results of detailed experimental investigation of the process of electrooptical switching illustrate limitations posed on the parameters of LC layers and switching fields.

2. NUMERICAL SIMULATION

The laboratory coordinate system *x*, *y*, *z* is oriented so that the *xy* plane coincides with one surface of a

plane-parallel LC layer and the z axis points toward the other surface of the layer. The LC layer is assumed to be homogeneous in the x and y directions. An important role in the simulation of bistable switching is played by the hydrodynamics. Based on the Frank–Oseen continuum theory of elasticity [12, 13], the Ericksen–Leslie hydrodynamic theory [14, 15], and the Euler–Lagrange formalism, it is possible to write the following system of dynamic equations for the LC layer with an electric field applied along the normal (coinciding with the z axis of the laboratory coordinate system), in which the local molecular orientation field is determined by the director $\mathbf{n} = (n_x, n_y, n_z) \equiv (\xi, \eta, \zeta)$:

$$\rho\frac{\partial v_x}{\partial t} = \frac{\partial}{\partial z}\Big[\alpha_2\zeta\frac{\partial\xi}{\partial t} + \alpha_3\xi\frac{\partial\zeta}{\partial t} + \frac{1}{2}\alpha_3\xi\eta v_y' + \frac{1}{2}(2\alpha_1\xi^2\zeta^2 - \alpha_2\zeta^2 + \alpha_3\xi^2 + \alpha_4 + \alpha_5\zeta^2 + \alpha_6\xi^2)v_x'\Big],$$

$$\rho\frac{\partial v_y}{\partial t} = \frac{\partial}{\partial z}\Big[\alpha_2\zeta\frac{\partial\eta}{\partial t} + \alpha_3\eta\frac{\partial\zeta}{\partial t} + \frac{1}{2}\alpha_3\xi\eta v_x' + \frac{1}{2}(2\alpha_1\eta^2\zeta^2 - \alpha_2\zeta^2 + \alpha_3\eta^2 + \alpha_4 + \alpha_5\zeta^2 + \alpha_6\eta^2)v_y'\Big], \quad (1)$$

$$I\frac{\partial^2\xi}{\partial t^2} = \lambda\xi - \frac{\partial F}{\partial\xi} + \frac{d}{dz}\frac{\partial F}{\partial\xi'} - \gamma\frac{\partial\xi}{\partial t} - \alpha_2\zeta v_x',$$

$$I\frac{\partial^2\eta}{\partial t^2} = \lambda\eta - \frac{\partial F}{\partial\eta} + \frac{d}{dz}\frac{\partial F}{\partial\eta'} - \gamma\frac{\partial\eta}{\partial t} - \alpha_2\zeta v_y',$$

$$I\frac{\partial^2\zeta}{\partial t^2} = \lambda\zeta - \frac{\partial F}{\partial\zeta} + \frac{d}{dz}\frac{\partial F}{\partial\zeta'} - \gamma\frac{\partial\zeta}{\partial t} - \alpha_3\xi v_x' - \alpha_3\eta v_y'.$$

Here, ρ is the LC density, I is the moment of inertia per unit LC volume, $\mathbf{v} = (v_x, v_y, v_z)$ is the velocity vector of the hydrodynamic flow, α_1–α_6 are the hydrodynamic Leslie coefficients, $\gamma = \alpha_3 - \alpha_2$ is the rotational viscosity, $\alpha_6 = \alpha_5 - \gamma$, λ is the Lagrange multiplier related to the unit normal vector $\mathbf{n}$, F is the free energy density of the LC layer given by the formula

$$F = \frac{1}{2}\{K_{11}\zeta'^2 + K_{22}(\eta\xi' - \xi\eta' + q_0)^2 + K_{33}[\zeta^2(\eta'^2 + \xi'^2) + (\xi\xi' + \eta\eta')^2]\} + \frac{1}{2}\frac{D_z^2}{\varepsilon_\perp\left(1 + \frac{\varepsilon_\parallel - \varepsilon_\perp}{\varepsilon_\perp}\zeta^2\right)}, \quad (2)$$

K_{ii} are the elastic moduli, and $\varepsilon_\parallel$, $\varepsilon_\perp$ are the principal components of the low-frequency permittivity tensor; primed symbols denote the derivatives of the components of director $\mathbf{n} = (n_x, n_y, n_z)$:

$$\frac{\partial n_x}{\partial z} \equiv \xi', \quad \frac{\partial n_y}{\partial z} \equiv \eta', \quad \frac{\partial n_z}{\partial z} \equiv \zeta'. \quad (3)$$

In expression (2) for the free energy density F, the contribution due to the electric field is written in terms of the z-component of the electric displacement vector $\mathbf{D}$. In other words, we adopt a model in which a charge (rather than a potential difference) is set on the LC boundaries (electrodes). For this reason, the term describing the electric field contribution enters into the free energy with the plus sign. Alternatively, the electric contribution is frequently included with the minus sign, which indicates that the work is performed by a voltage source and implies that a potential difference is set on the boundaries. Both models eventually lead to identical results. However, description in terms of the electric displacement is more convenient because LCs are good dielectrics and the electric displacement (in contrast to the electric field strength $\mathbf{E}$) is independent of z even for an inhomogeneous distribution of the director. The voltage drop across the LC layer and the z-component of the electric displacement are related as

$$V = \int_0^d E dz = \frac{D_z}{\varepsilon_\perp}\int_0^d\left(1 + \frac{\varepsilon_\parallel - \varepsilon_\perp}{\varepsilon_\perp}\zeta^2\right)^{-1} dz, \quad (4)$$

where d is the LC layer thickness.

It was experimentally established that the inertial contributions written in the left-hand sides of Eqs. (1) are negligibly small. These terms were ignored in the simulation. In solving Eqs. (1), we used the boundary conditions determined by a anchoring energy and an LC director pretilt angle at the alignment surfaces. The boundary-value problem was solved by a method analogous to that described in [16], using experimentally determined values of the director pretilt angle (4°) on both boundaries of the LC layer and a anchoring energy of $W = 0.2$ mJ/m^2. It is important to note, however, that the finiteness of the anchoring energy of LC molecules at the boundaries is not substantial for the effect under consideration. For this reason, Eqs. (1) can be also solved with rigid boundary conditions (infinitely high anchoring energy).

The simulation was performed for the following viscoelastic and dielectric parameters of the model LC:

$$K_{11} = 6 \text{ pN}, \quad K_{22} = 3 \text{ pN},$$

$$\alpha_1 = 0, \quad \alpha_2 = -0.083 \text{ Pa s}, \quad \alpha_3 = 0,$$

$$\alpha_4 = 0.075 \text{ Pa s}, \quad \alpha_5 = 0.1 \text{ Pa s},$$

$$\varepsilon_\parallel = 19, \quad \varepsilon_\perp = 6.$$

These values are close to the parameters of the well-known, widely used LC material 4-*n*-pentyl-4'-cyanobiphenyl (5CB) [17]. The system of equations (1) was numerically solved by the well-known relaxation method using the NLCL program package developed by one of the authors (S.P.P.) for the simulation of electrooptical effects in LCs. In addition, it was necessary to solve the corresponding optical problem for inhomogeneous anisotropic media. This solution was obtained using an algorithm described in detail elsewhere [18].

In expression (2), the term q_0 characterizes the chiral properties of the LC material. In the course of simulation, q_0 and the elastic modulus K_{33} were the variable parameters. In bulk materials, where the contribution of the anchoring energy to the total free energy of the system is negligibly small, the equilibrium LC texture with a minimum free energy corresponds to a helical director distribution with a natural pitch of $P_0 = 2\pi/q_0$ of the cholesteric helix. In the case of thin LC films, the situation dramatically changes. Owing to the surface anchoring, multiple local minima of the free energy appear that correspond to the states with various numbers of half-turns of the cholesteric helix accommodated within the layer thickness. The energy barrier between the adjacent states increases with the anchoring energy. Thus, in the rigid anchoring limit, the ground state with the pitch P_0 is accompanied by an infinite number of metastable states of the director differing by an integer number of half-turns of the helix over the LC layer thickness. Here, a substantial point is that a continuous transition between two states differing by an odd number of half-turns is impossible for a continuous deformation of the director in the entire volume. Therefore, such states are topologically nonequivalent and transitions between such states without defect formation are impossible. It is interesting to note that, in the case of symmetric boundary conditions (even for a small anchoring energy making possible a change in the state of director orientation at the boundaries), analogous nondefect transitions are also forbidden: such a transition would require the director to rotate by 90° in the middle of the LC layer, which contradicts the symmetry of the system [11].

The situation is different for the states differing by an even number of half-turns of the helix. In this case, there are no prohibitions and nondefect switching is possible. For example, let us consider the states schematically depicted in Fig. 1. Here, nondefect transitions from the state with one half-turn of the cholesteric helix over the layer thickness (180° twist state, Fig. 1a) to a unwound state (Fig. 1b) or to the state with one complete turn (Fig. 1c) are forbidden. For the sake of brevity, the states depicted in Figs. 1a–1c will be referred to below as π-, 0- and 2π-states, respectively. At the same time, transitions between the 2π- and 0-states are possible via a homeotropic state of the director in the middle of the layer (Fig. 1d). As was noted above, the possibility of such switching was demonstrated by Berreman and Heffner [5]. An important feature of the Berreman–Heffner system is that the ratio of the LC layer thickness (d) to the natural helix pitch (P_0) is $d/P_0 \approx 0.5$, which accounts for approximately equal energies of the 0- and 2π-states. However, this circumstance also implies that, for the planar boundary conditions, the minimum free energy corresponds to the π-state. Therefore, the realization of switching between the 2π- and 0-states requires an intermediate process to convert the ground π-state to the topologically nonequivalent 2π-state, which essentially involves the defect formation.

The results of our numerical calculations showed that the application of a driving voltage waveform (Fig. 2b) comprising a sequence of two-part and single pulses (similar to that used in [7]) makes it possible to switch the system between the 2π- and 0-states even if the 2π-state is the ground state. Figure 2a shows the simulated electrooptical response of an LC layer sandwiched between crossed polarizers. Here, the director passes from a twist state to the homeotropic state upon application of the first two-part voltage pulse to the LC layer, and exhibits relaxation to the homogeneous state when the field is switched off. The transition is accompanied by the characteristic bounce of the optical transmission. The final homogeneous 0-state is dark in the crossed polarizers because the axis of one polarizer and the LC director occur in the same plane. Arrival of the following single pulse switches the homogeneous 0-state back to the 2π-state which (for the given orientation of polarizers and optical anisotropy of the LC layer) renders the optical system transparent. Similarly to the case of the Berreman–Heffner effect, the transitions proceed via the homeotropic state of the director (aligned in the z axis) in the middle of the layer (Fig. 1d) owing to different development of the hydrodynamic flows dependent on the shape of the driving voltage.

The results of our simulations are summarized in Fig. 3. It was established that an important condition for bistable switching in the system studied is a large ratio of the elastic moduli: $K_{33}/K_{22} > 3$. Curves *1* and *2* in Fig. 3 separate regions representing various states of the director after the field is switched off, depending on the variable parameters K_{33}/K_{22} and d/P_0. In the region of K_{33}/K_{22} and d/P_0 ratios to the left of curve *1*, the field induces a homogeneous state which cannot be switched back to the 2π-state. To the right of curve *2*, the given driving voltage waveform induces a 2π-state without the possibility of switching to the homogeneous 0-state. And it is only in the region between curves *1* and *2* that the situation admits switching between the 0- and 2π-states. For $0.5 < d/P_0 < 0.75$, the system exhibits the Berreman–Heffner switching in spite of the thermodynamically favorable π-state. The system will relax into this third state via defect formation if the field is switched off for a prolonged time.

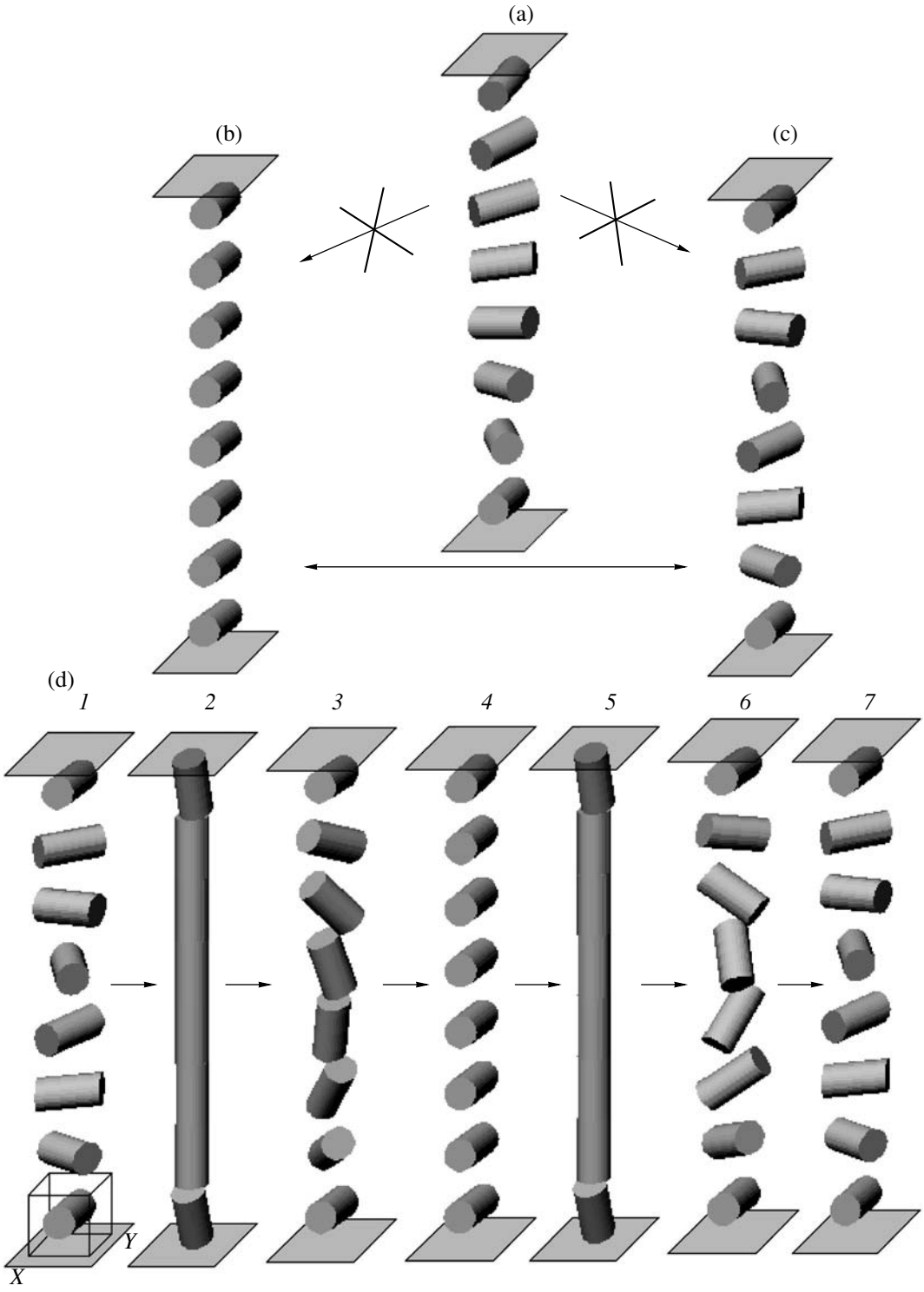


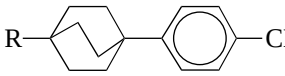
Fig. 1. Schematic diagram of the LC director distribution in depth of an LC layer: (a) helical 180° twist state (π-state); (b) homogeneous unwound state (0-state); (c) helical 360° twist state (2π-state); (d) the process of switching from the (*1*) 2π- to (*4*) 0-state and (*5*–*7*) the reverse process (states *2* and *5* are induced by an applied electric field).

Another important result revealed by the simulation is that, as can be seen from Fig. 3, the switching is also possible for $0.75 < d/P_0 < 1.25$, where the 2π-state becomes the ground state under strong anchoring conditions. Although the energy of the metastable 0-state increases so that the π-state becomes energetically more favorable than the 0-state, the system switches to the 0-state because it is topologically equivalent to the

2π-state. Thus, the third (π-) state in this regime is eliminated.

3. EXPERIMENT

In searching for the system featuring the predicted bistable switching, we selected LCs of the two types. The first LC is the aforementioned 5CB. This material is frequently used as the model LC medium, although the elastic moduli of 5CB-based LCs have been reported only in a few papers [17, 19]. According to the published data, the room-temperature ratio of the elastic moduli K_{33}/K_{22} (elastic ratio) for 5CB is close to three, which allowed us to expect bistable switching to take place even in this classical material. The second LC, referred to below as BCO, was prepared from a mixture of three bicyclooctane homologs with a common structural formula,



(5)

where R = C_3H_7, C_5H_{11}, and C_7H_{15}. The obtained mixture contained the three homologs in a weight ratio of 30 : 40 : 30. The physical properties of an analogous mixture have been thoroughly studied by Bradshaw *et al.* [20] and are presented in the table. It should be noted that the measurements reported in [20] were performed at relatively high temperatures; nevertheless, the material is characterized by an extremely high elastic ratio, $K_{33}/K_{22} = 3.3$, which must still increase (taking into account the temperature derivatives) on the passage to room temperature. The BCO-based LC mixture fully justified our expectations with respect to bistable switching in a system with the ground 2π-state according to the state diagram in Fig. 3.

The chiral additive was an KhDN-1 compound (synthesized at the State Research Institute of Organic Intermediates and Dyes (NIOPIK), Moscow), which was introduced into both LC matrices in a concentration of 0.4 wt %. This amount of additive ensured the natural helical pitch of $P_0 = 10.6$ μm for 5CB and $P_0 =$ 6.9 μm for BCO.

The experiments were performed on samples in the form of a Cano wedge with the layer thickness varying from 2 to 10 μm over a distance of about 10 cm. The wedge was obtained between polyimide-coated 10 × 15 × 100 mm optical glass plates with a pattern of 4-mm-wide transparent SnO_2 electrodes perpendicular to the long side, formed by chemical etching. The high rigidity of glass plates and high quality of their surfaces ensured smooth variation of the LC layer thickness. The LC layers were aligned by rubbing the polyimide films, which ensured the required planar boundary conditions with strong anchoring and a director pretilt angle of about 4°. Because of the strong anchoring of LC molecules at the alignment surfaces, the pitch cannot vary smoothly along the wedge. This results in the formation of the Grandjean zones in which the number of half-turns of the helix is fixed within a certain interval of the layer thickness, as deter-

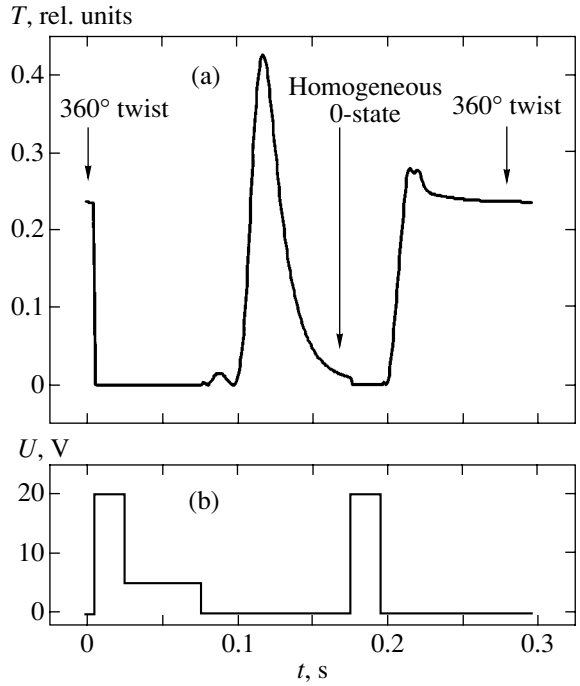


Fig. 2. Bistable switching in an LC layer: (a) numerically simulated electrooptical response showing transitions between 0- and 2π-states; (b) control pulsed voltage waveform. The LC layer is sandwiched between crossed polarizers; the transmission axis of one polarizer is parallel to the *xz* plane containing the LC director in the homogeneous 0-state. The numerical simulation was performed for $d/P_0 = 1$, $K_{33}/K_{22} = 4$, and $\Delta n = 0.19$ (other parameters of the virtual LC are close to the corresponding values for 5CB (see the text).

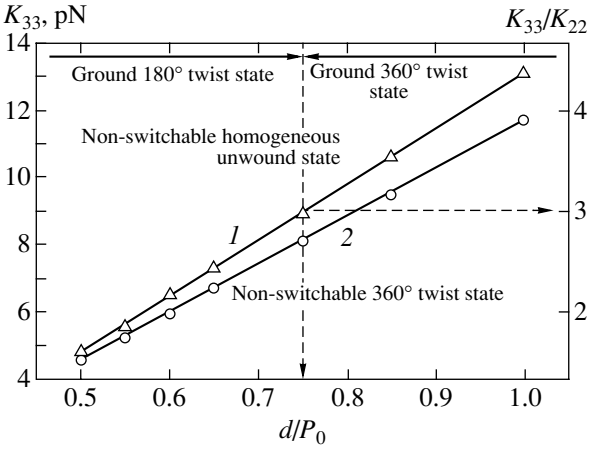


Fig. 3. The results of numerical simulations showing the regions of parameters of the virtual LC featuring non-switchable 0-state (to the left of curve *1*), bistable switching between the 2π- and 0-states (region between curves *1* and *2*), and non-switchable 2π-state (to the right of curve *2*) in response to a driving voltage waveform analogous to that depicted in Fig. 2b.

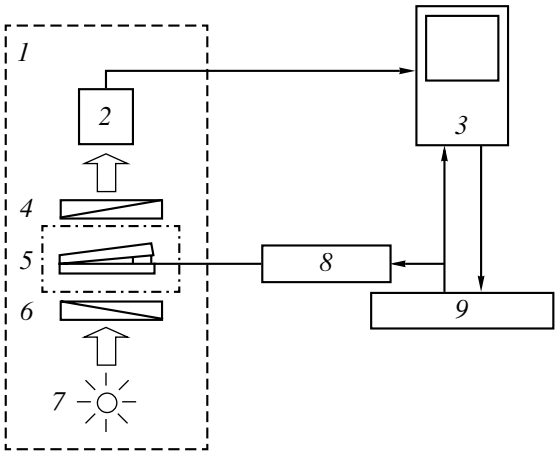


Fig. 4. Schematic diagram of the experimental setup for the investigation of LC orientation and electrooptical switching: (*1*) polarization microscope (Polam P-113); (*2*) photodetector or video camera; (*3*) digital oscillograph (Tektronix TDS 3012); (*4*, *6*) microscope polarizers; (*5*) sample (LC layer in Cano wedge); (*7*) microscope light source (incandescent lamp); (*8*) voltage source; (*9*) personal computer (with Creative PCI-128 sound board and PhysLab 5.0 program package).

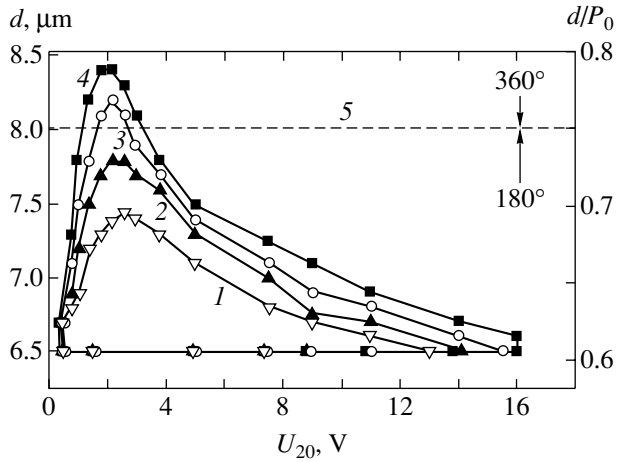


Fig. 5. Bistable switching in 5CB at $T = 21°C$: experimental curves bounding the region of values of the layer thickness d, the ratio d/P_0, and the selection pulse amplitude (U_{20}) ensuring bistable switching between the 2π- and 0-states for $U_{10} = 20$ V, $\tau_{10} = 50$ ms, and various selection pulse durations $\tau_{20} = 20$ (*1*), 50 (*2*), 100 (*3*), and 200 ms (*4*); horizontal line *5* corresponds to the LC layer thickness at which the disclination line separating the second (180° twist) and third (360° twist) Grandjean zones in the Cano wedge is observed.

mined by the d/P_0 ratio varying in the intervals 0–1/4, 1/4–3/4, 3/4–5/4 and so on. The numbers of half-turns of the cholesteric helix in the adjacent zones differ by one, so that the states of the LC director in the adjacent zones are topologically nonequivalent. The neighboring zones are separated by disclination lines.

The sample design described above makes it possible to study the electrooptical effects simultaneously in three Grandjean zones corresponding to a director twist angle of 0° (unwound or 0-state), 180° (π-state), and 360° (2π-state). Each of these zones consists of many local electrically independent cells of variable thickness. The local thickness of each cell was measured using a capacitive technique before filling the wedge with the LC medium.

Parameters of a BCO-based LC mixture [20]

Elastic moduli K_{11}, K_{22}, K_{33}, pN ($T = 74°C$)	11.2, 10.0, 33.0
dK_{11}/dT, pN/K ($T = 74$–$80°C$)	–0.3
dK_{33}/dT, pN/K ($T = 74$–$80°C$)	–1.3
$d(K_{22}/K_{11})/dT$, K^{-1} ($T = 74$–$80°C$)	–0.2
Principal components of the low-frequency permittivity tensor $\varepsilon_\parallel$, $\varepsilon_\perp$ ($T = 74°C$)	17, 8.5
Refractive indices $n_\parallel$, $n_\perp$ ($T = 74°C$)	1.64, 1.50
Bulk viscosity η, Pa s ($T = 20°C$)	0.1
NLC–isotropic phase transition temperature, °C	92
Crystal–nematic phase transition temperature, °C	20

The main idea in the search for and study of the predicted bistable switching consisted in monitoring the LC layer switching in the third (360° twist) Grandjean zone between the ground 2π-state and the homogeneous 0-state and vice versa. For a definite orientation of the axes of crossed polarizers, the 2π- and 0-states are readily distinguished by the naked eye due to the structure of differently colored bands separated by disclination lines. Indeed, if the axis of one of the crossed polarizers is oriented along the direction of rubbing of the alignment layers, the homogeneous state corresponds to the dark region irrespective of the LC layer thickness. Therefore, switching into the homogeneous state renders the interval of thicknesses corresponding to the third Grandjean zone optically equivalent to that of the first Grandjean zone. However, the third and first Grandjean zones are separated by a well-distinguished second Grandjean zone in which the director has a 180° twist.

The electrooptical response was studied using the setup schematically depicted in Fig. 4. The setup comprised a polarization microscope (equipped with a photodetector and a video camera), a digital oscillograph, and an original system of virtual devices PhysLab 5.0 used in this case for the formation of the control pulsed voltage waveform of a desired shape filled with a 1-kHz sinusoidal signal. Switching into the homogeneous 0-state was provided by a two-part pulse with the amplitudes U_{10} (reset phase) and U_{20} (selection phase) and the respective durations τ_{10} and τ_{20}. The reverse

switching to the 2π-state was induced by a single pulse with an amplitude of $U_{2\pi} = U_{10}$ (see Fig. 8b).

Figure 5 shows a family of closed curves bounding the region of values of the layer thickness (d/P_0) and the selection pulse amplitude (U_{20}) for bistable switching between the 2π- and 0-states in a 5CB-based LC for various selection phase durations τ_{20}. For example, at U_{20} = 2.5 V and τ_{20} = 200 ms, bistable switching is observed in the maximum range of LC thicknesses (from 6.5 to 8.5 μm). At a given content of the chiral additive, this corresponds to the interval of d/P_0 values from 0.61 to 0.8. For $d/P_0 < 0.75$, the switching is localized in the second Grandjean zone and corresponds to the Berreman–Heffner type bistability. Observation of the bistable switching effect in this region required multiply repeated application of the pulsed voltage waveform with a high amplitude (about 20 V), which was necessary to eliminate the ground π-state. In this case, we have initially observed in the Cano wedge a slow (over several seconds) transition from the π- to 2π-state via the formation of numerous defects. It was not until complete disappearance of the LC regions in the π-state that homogeneous bistable switching between the 2π- and 0-states could be observed. As for a very narrow interval within $0.75 < d/P_0 < 0.8$, this region falls into the third Grandjean zone. Here, the 2π-state is (by definition) the ground state and, hence, we observe the anticipated bistability without the third π-state. It should be emphasized that this thickness region of switchable states is very small with respect to the LC layer thickness and is localized at the very boundary of the second Grandjean zone. A comparison of the experimental data to the results of simulation showed that the elastic ratio K_{33}/K_{22} = 3 for the 5CB-based LC is somewhat overstated; nevertheless, to within the experimental uncertainty, this value is in good agreement with the observed behavior. Indeed, according to the results of simulation (Fig. 3), the upper boundary of the region of switchable states with respect to the LC layer thickness corresponds to d/P_0 = 0.82, while the experiment (for the optimum parameters of switching pulses) gives upper value of d/P_0 = 0.8.

The experimental situation for the BCO-based LC was sharply different from that for 5CB. As expected based on the simulation results for a large elastic ratio ($K_{33}/K_{22} > 3.3$), the switching was observed only in the third Grandjean zone. The observed behavior is illustrated in Fig. 6, which shows (like Fig. 5 for 5CB) a family of closed curves bounding the region of parameters ensuring bistable switching for various values of τ_{20}. At U_{20} = 5 V, an increase in the selection pulse duration τ_{20} from 20 to 200 ms results in that an increased (from 30 to 70%) part of the third zone is involved in the switching process. Figure 7 illustrates the role of the reset pulse amplitude U_{10} at a fixed reset time τ_{10} = 50 ms. As can be seen, the maximum range of layer thicknesses for bistable switching is achieved at relatively high U_{10} values on a level of 20–30 V. In the case of relatively low control fields corresponding to U_{10} = 10 V and U_{20} = 5 V, it is necessary to use

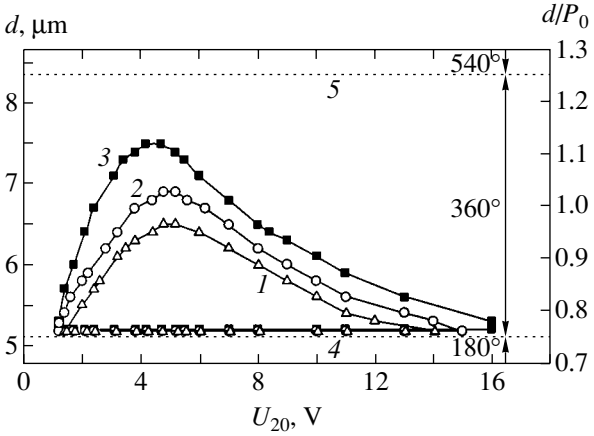


Fig. 6. Bistable switching in BCO at T = 24°C: experimental curves bounding the region of values of the layer thickness d, the ratio d/P_0, and the selection pulse amplitude (U_{20}) ensuring bistable switching between the 2π- and 0-states for U_{10} = 20 V, τ_{10} = 50 ms, and various selection pulse durations τ_{20} = 20 (*1*), 50 (*2*), and 200 ms (*3*); horizontal lines corresponds to the LC layer thicknesses at which the disclination lines separating (*4*) the second (180° twist) and third (360° twist) Grandjean zones and (*5*) the third and fourth (540° twist) Grandjean zones are observed.

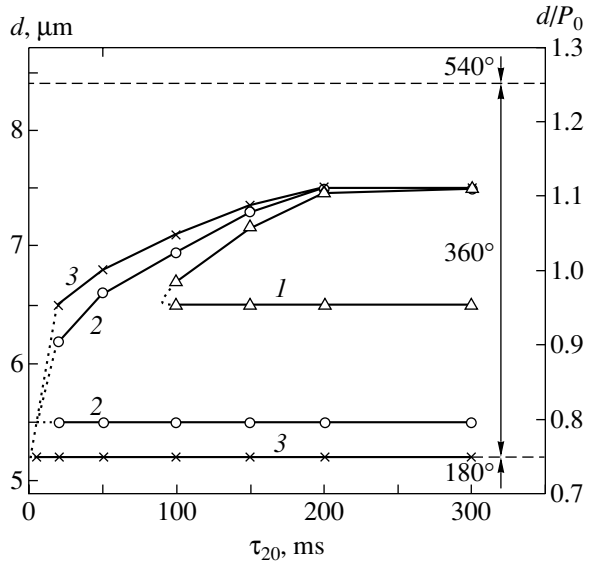


Fig. 7. Bistable switching in BCO at T = 24°C: experimental curves bounding the region of values of the layer thickness d, the ratio d/P_0, and the selection phase duration (τ_{20}) ensuring bistable switching between the 2π- and 0-states for U_{20} = 5.2 V, τ_{10} = 50 ms, and various reset pulse amplitudes U_{10} = 10 (*1*), 15 (*2*), and 30 V (*3*); dashed horizontal lines corresponds to the LC layer thicknesses at which the disclination lines separating the second (180° twist) and third (360° twist) Grandjean zones (for d/P_0 = 0.75) and the third and fourth (540° twist) Grandjean zones (for d/P_0 = 1.25) are observed.

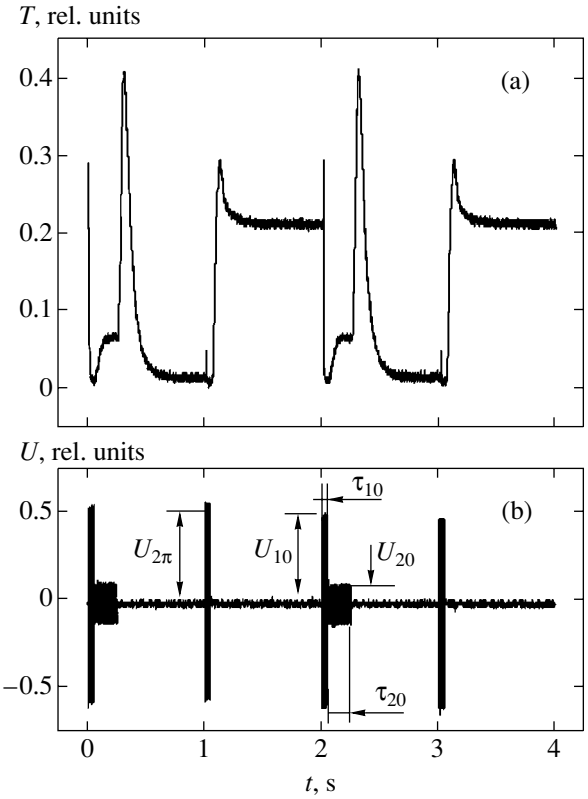


Fig. 8. Oscillograms of the (a) optical transmission T and (b) control pulsed voltage waveform for bistable switching between the 2π- and 0-states in the third Grandjean zone ($d/P_0 = 0.94$) at $T = 25°C$. The LC layer is sandwiched between crossed polarizers; the dark state (e.g., at $t = 1$ s) corresponds to the 0-state; the bright state (e.g., at $t = 1.8$ s) corresponds to a helical distribution of the director in the 2π-state.

relatively large durations of the selection phase: $\tau_{20} > 150$ ms. In this case, bistable switching is localized at the center of the third Grandjean zone where $d/P_0 \sim 1$.

Figure 8 presents oscillograms illustrating the electrooptical switching between the 0- and 2π-states. As can be seen, the observed electrooptical response coincides in detail with the predicted behavior (see Fig. 2). The difference in the time scale between Figs. 2 and 8 is related only to the difference in viscoelastic parameters between the model and real LC materials. It should be noted that a transition to the homogeneous 0-state after termination of the two-part voltage pulse is accompanied by the characteristic bounce of the optical transmission. The duration of this bounce is determined by viscoelastic properties of the LC medium. From the standpoint of practical applications, this optical bounce is undesired. The problem of optical compensation for this effect is a subject for separate investigation.

Finally, it should be noted that, since the unwound homogeneous state is metastable, prolonged absence of the control field leads to unavoidable relaxation to the ground 2π-state. The rate of this relaxation decreases with increasing LC layer thickness and with decreasing number of defects on the alignment surface. At a layer thickness of about 7 μm, the characteristic time of the complete transition to the ground state amounts to tens of seconds. In this context, it would be more correct to speak of a quasi-bistable switching. We use the term bistable switching because it is historically attributed to the Berreman–Heffner effect (which would be also more correctly classified as quasi-bistable switching).

4. CONCLUSIONS

We have theoretically predicted, based on the results of numerical simulation, and experimentally observed and studied the phenomenon of bistable switching in the third Grandjean zone of a chiral nematic layer. In this case, the ground (thermodynamically equilibrium) state corresponds to a helical director distribution deep in the layer with the resulting 360° twist, and the switching takes place between the ground state and a metastable unwound state of the director. In contrast to the well-known Berreman–Heffner bistable switching, the process under consideration does not involve a third state corresponding to the 180° twist. The results provide a deeper insight into the role of topologically equivalent states in the electrooptics of LCs. The absence of the topologically nonequivalent 180° twist state leading to undesired defect formation simplifies the use of the proposed effect in electrooptical devices.

ACKNOWLEDGMENTS

The authors are grateful to L.M. Blinov, V.V. Lazarev, B.A. Umanskiĭ, and the staff of the Laboratory of Liquid Crystals at the Shubnikov Institute of Crystallography for fruitful discussions and useful advice.

This work was supported by the Russian Foundation for Basic Research (project nos. 03-02-17288 and 04-02-16466) and the program "New Materials and Structures" of the Department of Physics of the Russian Academy of Sciences. One of the authors (S.P.P.) gratefully acknowledges the support from the Foundation for Support of Russian Science.

REFERENCES

1. J.-H. Kim, M. Yoneya, and H. Yokoyama, Nature **420**, 159 (2002).
2. A. J. Davidson and N. J. Mottram, Phys. Rev. E **65**, 051710 (2002).
3. R. Barberi, M. Giocondo, and G. Durand, Appl. Phys. Lett. **60**, 1085 (1992).
4. C. Joubert, J. Angele, A. Boissier, *et al.*, in *SID 02 Digest* (2002), p. 30.
5. D. W. Berreman and W. R. Heffner, Appl. Phys. Lett. **37**, 109 (1980).

6. D. W. Berreman and W. R. Heffner, J. Appl. Phys. **52**, 3032 (1981).
7. C. D. Hoke, J. Li, J. R. Kelly, and P. J. Bos, Jpn. J. Appl. Phys. **36**, L227 (1997).
8. T. Z. Qian, Z. L. Xie, H. S. Kwok, and P. Sheng, Appl. Phys. Lett. **71**, 596 (1997).
9. G. D. Lee, G. H. Kim, K. H. Park, *et al.*, in *SID 99 Digest* (1999), p. 862.
10. J. Birgerson, J. Osterman, and K. Skarp, *Eurodisplay* (2002), p. 401.
11. S. P. Palto, Zh. Éksp. Teor. Fiz. **121**, 308 (2002) [JETP **94**, 260 (2002)].
12. C. W. Oseen, Ark. Mat., Astron. Fys. **19A**, 1 (1925).
13. F. C. Frank, Discuss. Faraday Soc. **25**, 19 (1958).
14. F. M. Leslie, Arch. Ration. Mech. Anal. **28**, 265 (1968).
15. J. L. Ericksen, Trans. Soc. Rheol. **5**, 23 (1961).
16. S. P. Palto, Kristallografiya **48**, 130 (2003) [Crystallogr. Rep. **48**, 124 (2003)].
17. L. M. Blinov and V. G. Chigrinov, *Electrooptic Effects in Liquid Crystal Materials* (Springer, New York, 1994).
18. S. P. Palto, Zh. Éksp. Teor. Fiz. **119**, 638 (2001) [JETP **92**, 552 (2001)].
19. J. D. Bunning, T. E. Faber, and P. L. Sherrell, J. Phys. (Paris) **42**, 1175 (1981).
20. M. J. Bradshaw, D. G. McDonnell, and E. P. Raynes, Mol. Cryst. Liq. Cryst. **70**, 289 (1981).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 100, No. 1, 2005, pp. 208–217.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 1, 2005, pp. 230–241.

STATISTICAL, NONLINEAR, AND SOFT MATTER PHYSICS

Transitions to Chaos in Intracavity Parametric Wave Mixing

T. V. Dmitrieva and N. M. Ryskin
Saratov State University, Saratov, 410012 Russia
e-mail: RyskinNM@info.sgu.ru
Received February 16, 2004

Abstract—The complex dynamics of intracavity three-wave mixing are discussed. Detailed results obtained by numerical simulation of routes to chaos are presented for a wide range of parameters. As the pump intensity increases, a complex sequence of alternating regular and chaotic self-modulation regimes is observed. The relationship between these regimes and soliton formation and propagation is analyzed. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION AND STATEMENT OF THE PROBLEM

Spatiotemporal chaos in nonequilibrium systems is one of the most challenging problems in modern nonlinear physics [1–4]. One important example of wave interaction leading to the onset of chaos is three-wave mixing, which manifests itself in fluid dynamics, plasma physics, radio engineering, electronics, nonlinear optics, and other areas [2–7].

In this paper, we analyze the nonlinear dynamics of intracavity three-wave mixing in a quadratic nonlinear medium. This system can be considered as a model of an optical parametric oscillator (OPO), as well as of other electromagnetic oscillators [8, 9]. Another example is acousto-optic interaction in a system with reflecting boundaries [10]. Currently, generation of chaotic signals is considered a promising field of research in view of possible applications in communication systems, radar technology, etc. [11–13]. In these applications, distributed systems offer certain advantages as sources of high-dimensional chaotic signals characterized by a high complexity (in a certain sense).

We consider three interacting waves whose frequencies and wavenumbers satisfy the following phase matching conditions:

$$\begin{aligned} \omega_s + \omega_i &= \omega_p, \\ k_s + k_i &= k_p. \end{aligned} \tag{1}$$

A pump photon with frequency ω_p is down-converted into a signal–idler photon pair with frequencies ω_s and ω_i, respectively [2–7]. We use the model of this system developed in [14] by applying the slowly varying envelope approximation [2–6]:

$$\frac{\partial A_s}{\partial t} + v_s\frac{\partial A_s}{\partial x} = \sigma_s A_i^* A_p, \tag{2}$$

$$\frac{\partial A_i}{\partial t} + v_i\frac{\partial A_i}{\partial x} = \sigma_i A_s^* A_p, \tag{3}$$

$$\frac{\partial A_p}{\partial t} + v_p\frac{\partial A_p}{\partial x} = -\sigma_p A_s A_i, \tag{4}$$

where $A_{s,i,p}$, $v_{s,i,p}$, and $\sigma_{s,i,p}$ denote the corresponding amplitudes, group velocities, and nonlinear coupling constants, respectively. The nonlinear medium is confined in a cavity tuned to the signal frequency. The cavity is supposed to be transparent for the idler and pump beams. (This system is known as singly resonant optical parametric oscillator in the OPO theory [9].) Boundary conditions are set as follows (see [14]):

$$A_s(0, t) = RA_s(l, t - \Delta t), \tag{5}$$

$$A_i(0, t) = 0, \quad A_p(0, t) = A_0, \tag{6}$$

where $R = \rho\exp(i\psi)$ is a (complex-valued) feedback parameter, l is the cavity length, Δt is the feedback delay time, and A_0 is a constant input pump amplitude (controlled by an external pump source). While the initial condition for the pump is $A_p(x, 0) = A_0$, both signal and idler are seeded by noise.

The parameters ρ, ψ, and Δt depend on the cavity type, which is not specified here. For example, the forward- and reflected-wave amplitudes at the boundaries

of a linear Fabry–Perot cavity, A_s and A_s^-, satisfy the relations

$$A_s(0) = R_0 A_s^-(0),$$
$$A_s^-(l) = R_l A_s(l)\exp(-2ik_s l), \quad (7)$$

where R_0 and R_l denote the (complex-valued) left and right boundary reflection coefficients, respectively. Since the reflected wave is not involved in the wave mixing, A_s^- obeys the equation

$$\frac{\partial A_s^-}{\partial t} - v_s\frac{\partial A_s^-}{\partial x} = 0 \quad (8)$$

for a wave propagating backwards with the signal group velocity. Representing the solution to Eq. (8) as

$$A_s^-(0, t) = A_s^-(l, t - l/v_s),$$

we eliminate the reflected-wave amplitude from (7) and obtain Eq. (5) with $\rho = |R_0R_l|$, $\psi = \arg(R_0R_l) - 2k_sl$, and $\Delta t = l/v_s$.

In the case of a traveling-wave ring cavity, the delay Δt is an independent parameter, but boundary conditions (5) and (6) still hold. Moreover, they apply to the system considered in [10], where no cavity was employed and feedback was implemented by means of acoustic wave reflection from the boundaries.

Changing to the dimensionless variables (see [14])

$$A_s' = \sqrt{\frac{\sigma_p v_s}{\sigma_s v_p}}\frac{A_s}{A_0}, \quad A_i' = \sqrt{\frac{\sigma_p v_i}{\sigma_i v_p}}\frac{A_i}{A_0}, \quad A_p' = \frac{A_p}{A_0},$$
$$\xi = \frac{x}{l}, \quad \tau = \frac{v_s v_i(t - x/v_i)}{l(v_i - v_s)},$$

and dropping the primes, we rewrite Eqs. (2)–(4) as

$$\frac{\partial A_s}{\partial \tau} + \frac{\partial A_s}{\partial \xi} = \alpha A_i^* A_p, \quad (9)$$

$$\frac{\partial A_i}{\partial \xi} = \alpha A_s^* A_p, \quad (10)$$

$$(1 + u)\frac{\partial A_p}{\partial \tau} + \frac{\partial A_p}{\partial \xi} = -\alpha A_s A_i, \quad (11)$$

where $\alpha = A_0 l(\sigma_s\sigma_i/v_s v_i)^{1/2}$ is the normalized input pump amplitude and the parameter $u = v_i(v_s - v_p)/v_p(v_i - v_s)$ characterizes the group-velocity mismatch. To be specific, we assume that $v_i > v_s$; otherwise, similar equations can easily be obtained by changing the normalization conditions. In either case, the difference of the signal and pump group velocities is supposed to be sufficiently large, because their relative values determine the parametric-gain bandwidth. When $v_s \approx v_i$, the bandwidth tends to infinity, i.e., the corresponding signal spectrum is too wide to treat the parameters in the starting equations as constants, and even the equations for slowly varying amplitudes are not valid anymore.

Boundary conditions (5) and (6) rewritten in terms of the dimensionless variables are

$$A_s(0, \tau) = RA_s(1, \tau - \delta), \quad A_i(0, \tau) = 0,$$
$$A_p(0, \tau) = 1, \quad (12)$$

where

$$\delta = \frac{v_s v_i(\Delta t + l/v_i)}{l(v_i - v_s)}$$

is the normalized delay. For the Fabry–Perot cavity, $\Delta t = l/v_s$ (see above) and, therefore,

$$\delta = \frac{v_i + v_s}{v_i - v_s} > 1.$$

However, δ can have any value in the present analysis, since it is not restricted to any specific type of cavity. Under the present assumptions, the spectrum of the system is given by the expression (see [14])

$$\Omega_n = \frac{2\pi n + \psi}{1 + \delta}, \quad (13)$$

where Ω is a dimensionless offset relative to the frequency ω_s satisfying exact phase matching conditions (1).

It was shown in [14] that the pump amplitude corresponding to the generation threshold is a periodic function of the feedback-parameter phase ψ. This behavior is common to delayed feedback systems (e.g., see [15–18]). When $\psi = -2\pi n$ ($n = 0, 1, 2, \ldots$), the offset given by (13) is zero (i.e., one of the dimensional eigenfrequencies corresponds to the exact phase matching conditions) and generation is initiated by a pump of minimal intensity. In this case, the generation condition has a simple form:

$$\rho\cosh\alpha = 1. \quad (14)$$

Conversely, when $\psi = -2\pi n + \pi$, the nearest eigenfrequencies are equally separated from the phase-matching frequency, the threshold pump intensity is at its highest, and there is a periodic array of generation zones in the (α, ψ) plane.

As the pump intensity increases, the steady-state generation regime (with time-independent envelopes) becomes unstable. In the ensuing self-mode-locked

regime, a parametric soliton periodically propagates along the axis [14, 15]. The onset of the self-mode-locked regime always occurs via a "hard" change from steady-state generation to strong envelope oscillation. This can be interpreted as simultaneous generation of many phase-locked cavity modes leading to periodic formation of soliton-like pulses. This self-mode-locking mechanism is associated with a dip in the gain curve [14, 17–21].

According to [14, 15], a transition to chaos through intermittency is observed at the center of a generation zone as the phase locking characteristic of the steady-state and self-mode-locked periodic regimes breaks down. In the latter regime, the phases of the generated modes (each depending separately on coordinate and time) satisfy the relation

$$\sin[\varphi_p(\xi, \tau) - \varphi_s(\xi, \tau) - \varphi_i(\xi, \tau)] = 0, \qquad (15)$$

which ensures optimal conditions for energy exchange between the modes. As the modulation depth increases with α, the system passes through states with fast-varying phases. At certain instants, the phase locking breaks down. Accordingly, both coupling efficiency and envelope oscillation weaken, the phase locking is restored, and the process repeats. The output waveform is a sequence of almost regular intervals of phase-locked oscillation with growing modulation depth interrupted by intervals of irregular oscillation.

However, the description given in [14, 15] does not embrace all scenarios. Nonlinear oscillations in distributed systems are characterized by complex behavior in the corresponding parameter spaces involving alternate regular and chaotic regimes and different scenarios of transition to chaos [15–18, 22, 23]. The system considered here should be expected to exhibit behavior of this kind as well. Note also that only the transition to chaos at the center of a generation zone has been analyzed, whereas essentially different dynamics may be observed when the system is detuned from the zone center frequency [15, 16].

In this paper, we present the results of a detailed study of chaotic dynamics. In Section 2, we assume that phase-locking condition (15) is satisfied, in which case the envelope amplitudes in Eqs. (9)–(12) can be treated as real variables. We demonstrate a variety of self-mode-locked regimes differing by the number and polarity of the generated solitons. As the pump intensity increases, a transition to chaos occurs via breakdown of quasiperiodic dynamics. It was shown in [14] that the scenario described above cannot be implemented when the pump intensity substantially exceeds the generation threshold, because the phase-locked regime becomes unstable. However, an analysis of this simplified route provides a correct physical interpretation of the complex dynamics observed in the general case of complex amplitudes (see Section 3). In Section 4, we discuss a scenario of transition to chaos near the boundaries of a generation zone, where bistability is possible; i.e., the regimes developing from two neighboring modes coexist. The coalescence of these regimes (disappearance of bistability) due to increasing supercriticality is analyzed in detail.

It should be mentioned here that various issues arising in the OPO theory have been addressed in numerous studies (e.g., see [24–34]). However, the complex dynamics and scenarios of transition to chaos discussed in this paper have almost never been considered. The analyses were focused on stability of the steady-state regime, formation of soliton-like localized structures, and related phenomena. For the most part, the case when all of the three waves are phase-matched with high-Q cavity modes was considered (three-wave parametric oscillator). Under this condition, one can set periodic boundary conditions and seek solutions for the slowly varying envelope fields of appropriate eigenmodes (i.e., consider the so-called mean-field limit [24–26, 28, 29, 31, 34]). In this approximation, the starting equations can be reduced to models analogous to the Ginzburg–Landau or Swift–Hohenberg equation [25, 26, 28, 31]. However, the three-wave coupling model based on Eqs. (9)–(11) is better suited for analyzing the essentially multimode regimes of complex dynamics. According to the terminology of the OPO theory, this model describes a singly resonant nondegenerate OPO with an arbitrary-Q cavity. However, we emphasize that the model is universal in the sense that it describes the dynamics of intracavity parametric mixing of waves of any nature.

We also note that our analysis is restricted to a one-dimensional problem formulation. This approach is applicable when the transverse field distribution can be treated as invariant. Thus, we ignore the instabilities of transverse field profile, which play an important role in the dynamics of broad-area cavities [24–26, 28, 29, 32]. The complex dynamics due to two-dimensional effects should be analyzed in a separate study.

2. TRANSITION TO CHAOS AT THE CENTER OF A GENERATION ZONE: PHASE LOCKING

Consider the dynamics of three-wave mixing at the center of a generation zone ($\psi = 0$). In this case, it can be shown that there exist solutions describing phase-locked modes, i.e., satisfying condition (15) (see [14]), and the envelope in Eqs. (9)–(12) can be treated as real variables.

To be specific, we set $\rho = 0.2$ (i.e., consider a relatively low-Q cavity), $u = -1$, and $\delta = 1.0$. We consider the sequence of bifurcations observed as the input pump amplitude is increased. When α is relatively small, we are dealing with single-mode generation at the fundamental frequency (see [14]). Since $\Omega = 0$ in our model, the generation frequency is ω_s. When α exceeds a certain threshold α_{sm}, the instability of the

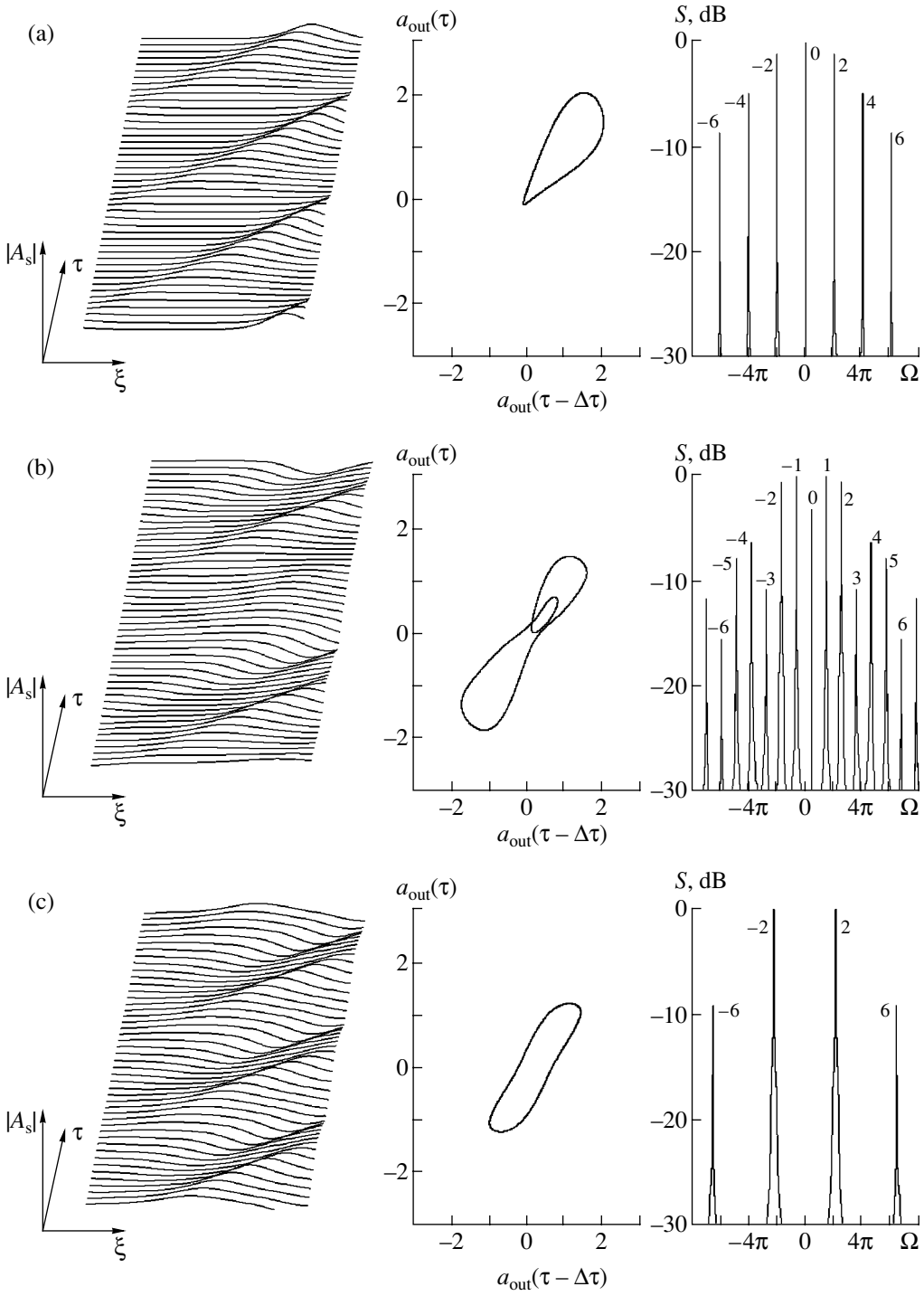


Fig. 1. Spatiotemporal dynamics, phase portraits, and spectra of output signal in different self-mode-locked regimes: $\psi = 0$; $u = -1$; $\delta = 1.0$; $\rho = 0.2$; $\alpha = 4.75$ (a), 5.0 (b), and 5.8 (c).

single-mode regime leads to a self-mode-locked regime. The output waveform $a_{out}(\tau) = |A_s(\xi = 1, \tau)|$, phase portrait, and spectrum shown in Fig. 1a illustrate the spatiotemporal dynamics of periodic soliton-like pulse propagation. The hard onset of self-mode-locking is characterized by generation of numerous side modes at frequencies close to Ω_n defined by (13). The particular combination of the side modes is determined by the value of δ (see [14] for details). In the case considered here, even modes are generated ($n = 2, 4, 6, \ldots$) and the offset Ω_{sm} between the fundamental mode and the nearest side modes is approximately 2π. Note that the self-mode-locking mechanism described here is characteristic of broadband delayed-feedback oscillators, such

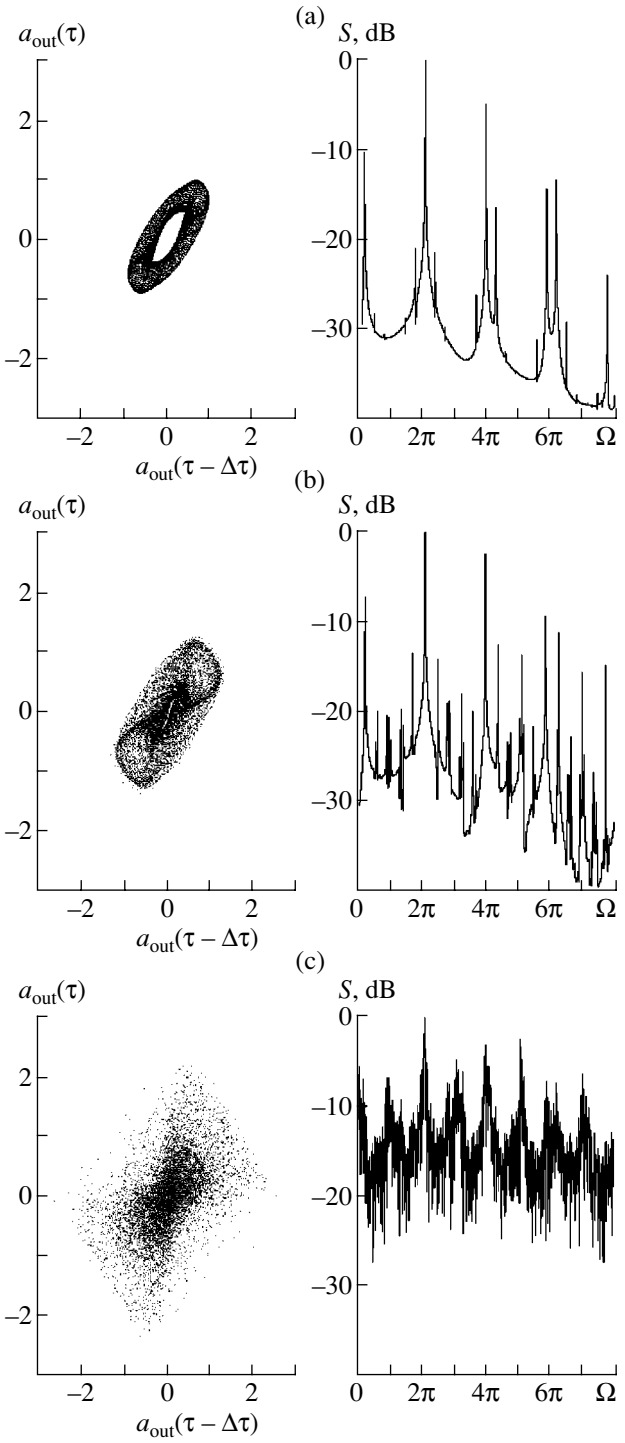


Fig. 2. Phase portraits and spectra illustrating a quasiperiodic route to chaos at $\alpha = 8.2$ (a), 8.5 (b), and 8.75 (c). The remaining parameters are as in Fig. 1.

as traveling-wave tube oscillators [18–21], ring-cavity lasers [35–37], and free electron lasers [38].

New self-mode-locked regimes emerge as α increases, differing by the number and polarity of generated parametric solitons. Figure 1 illustrates the three regimes of this kind obtained for the parameter values specified above. In the regime corresponding to Fig. 1a, periodic propagation of a single soliton with positive polarity is observed. Note that the obvious invariance of Eqs. (9)–(11) under the change $A_{s,i} \longrightarrow -A_{s,i}$ implies that there exists either a solution that is symmetric under this change or a pair of asymmetric solutions that are mutually symmetric. Figure 1a illustrates one of the admissible asymmetric regimes.

Figure 1b illustrates an asymmetric two-hump soliton regime with opposite polarities. The corresponding modulation frequency is half that in the case of Fig. 1a; i.e., odd side modes are generated. (Recall that the onset of a self-mode-locked regime observed when $\delta = 1.0$ is due to generation of the modes with $n = \pm 2$.) The section of the corresponding phase diagram shows a limit cycle of complex geometry. Note that the modes with $n = \pm 1$ have the largest amplitudes in this, whereas the largest amplitude in the regime illustrated by Fig. 1a corresponds to the fundamental mode ($n = 0$). The dip at the fundamental frequency manifests a change in the gain curve [14].

Figure 1c illustrates a symmetric regime in which single-hump solitons having similar shapes and opposite polarities (sometimes called solitons and antisolitons) are generated. In this case, the modulation frequency Ω_{sm} is close to 2π (as in Fig. 1a), but the fundamental mode is completely suppressed. The spectrum consists of the Ω_{sm}-mode and its odd harmonics; i.e., only the $\pm 2(n+1)$th modes survive ($n = 0, 1, 2, \ldots$).

Analogous transitions between asymmetric and symmetric self-mode-locked regimes have also been found in various models of optical lasers [35, 36, 39] (this phenomenon was called mode splitting in [39]). These results were indirectly confirmed in a recent study [40] of parametric spin-wave generation in an "active ring" consisting of an yttrium–iron–garnet film and a feedback loop containing a microwave amplifier and a variable attenuator. Both nonlinearity and dispersion were completely determined by the film's properties, while the amplifier was used to compensate for coupling losses in the input and output antennas. The film was pumped by a square-pulse train at the doubled signal frequency. When the signal frequency was close to one of the ring eigenfrequencies, a periodic sequence of identical solitons analogous to that shown in Fig. 1a was generated. When the signal frequency was approximately at the midpoint between eigenfrequencies, solitons with alternating polarity were generated in a sequence analogous to that presented in Fig. 1c. Similar behavior was exhibited by different active rings containing ferromagnetic films [41, 42]. However, the transitions between regimes predicted by the present model are due to variation of pump intensity rather than frequency and are associated with the nonlinear transformation of the gain curve described in [14].

With further increase in α, quasiperiodic self-modulation is observed, and then a transition to chaos

occurs via breakdown of quasiperiodic dynamics (the Ruelle–Takens scenario). Figure 2 shows phase portraits and spectra illustrating quasiperiodic and chaotic self-modulation regimes.[1]

Qualitatively similar behavior is observed at different values of ρ and δ. The influence of the feedback parameter is clear from the diagram of regimes in the (ρ, α) plane shown in Fig. 3. The curves shown here represent the generation threshold (predicted by (14)) and successive transitions from single-mode generation via self-mode-locked periodic regimes with a different number of solitons to quasiperiodic and chaotic self-modulation. The illustrated behavior deviates from the scenario described above only when ρ is close to unity (i.e., for high-Q cavities), in which case a hard transition to chaos is observed instead of quasiperiodic self-modulation. Note that the regions of quasiperiodic dynamics contain multiple synchronization tongues representing phase-locked regimes characterized by rational quotients of mode frequencies. They are not shown in the figure because they are too small.

Variation of the normalized delay δ does not lead to any qualitative change either. However, the overall dynamics becomes increasingly complex as this parameter increases. In particular, the number of multisoliton regimes observed prior to the onset of quasiperiodic self-modulation increases. This is easy to explain, because transitions between different regimes are due to competition between modes (see above), and the parameter δ controls the number of "active" modes (whose frequencies lie within the parametric-gain bandwidth). In any event, the typically observed phenomena include mode splitting and quasiperiodic route to chaos.

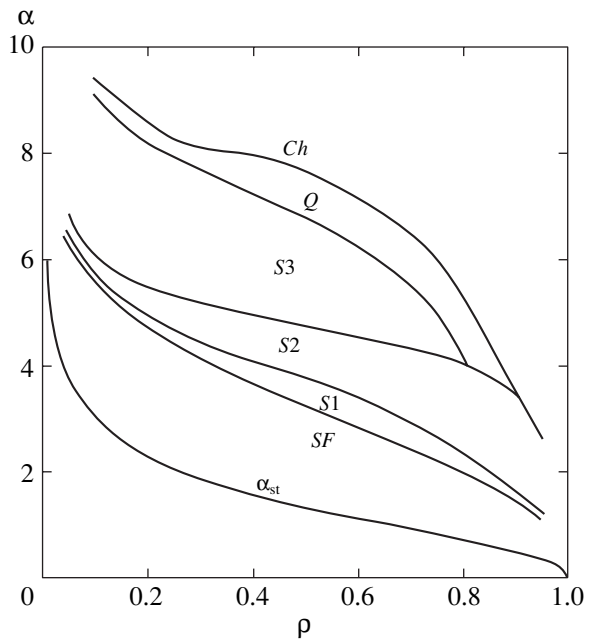


Fig. 3. Map of dynamic regimes in the (ρ, α) plane: α_{st} = generation threshold; *SF* = steady-state oscillation at the fundamental frequency; *S*1–*S*3 = self-mode-locked regimes with different number of generated solitons; *Q* = quasiperiodic self-modulation; *Ch* = chaotic self-modulation.

3. TRANSITION TO CHAOS AT THE CENTER OF A GENERATION ZONE: GENERAL CASE

Even though phase-locked modes always exist at the center of a generation zone, they may be unstable. Calculations performed in the general case of complex amplitudes show that the scenario described in the preceding section holds only for steady-state generation and self-mode-locked periodic regimes, when a single soliton forms over the cavity length (regions *SF* and *S*1 in Fig. 3) [14, 15]. When the formation of two-hump solitons with opposite polarities becomes possible, there exist points where the signal amplitude vanishes. Under this conditions, the system passes through states with fast-varying phases, phase-locking condition (15) is violated, and therefore both coupling efficiency and envelope oscillation weaken. Then, the mode locking is restored. Overall, the signal exhibits transition via intermittency to a chaotic regime characterized by a continuous spectrum. In the (ρ, α) plane shown in Fig. 3, the intermittency threshold corresponds to the boundary separating regions *S*1 and *S*2. Since condition (15) does not hold in chaotic regimes, the coupled wave amplitudes cannot be treated as real quantities and the multisoliton regimes described in the preceding section are, strictly speaking, unstable. However, since the time scales of transient processes are relatively long, the regimes should rather be considered as metastable. Therefore, they can be implemented in experiments, for example, by pumping with a pulse train (as was done in [40]).

However, intermittent chaotic regimes correspond to a very narrow interval of α. At higher α, a periodic self-modulation largely analogous to that illustrated by Fig. 1b is observed, but the corresponding spectrum is asymmetric (mode frequencies are shifted relative to the their "zero-point" values given by (13)). Further increase in pump amplitude leads to a transition to chaos that has much in common with that described in Section 2. As a typical example, consider the case of ρ = 0.7 and ψ = 0.1π (Fig. 4). When α = 3.9, quasiperiodic self-modulation is observed (see Fig. 4a), the spectrum is dominated by modes with $n = -2$ and +1, while the central modes (with $n = -1$ and 0) are suppressed. The corresponding spatiotemporal dynamics resemble the two-soliton regime depicted in Fig. 1b. At a higher pump amplitude, a transition to chaos occurs via breakdown of quasiperiodic dynamics (see phase portrait and spectrum in Fig. 4b). When α = 5.0 (Fig. 4c), periodic mode locking is restored, and regular oscillation similar

[1] Only the spectral domain of Ω > 0 is shown in Fig. 2, since the spectra corresponding to real variables are symmetric about the zero offset frequency.

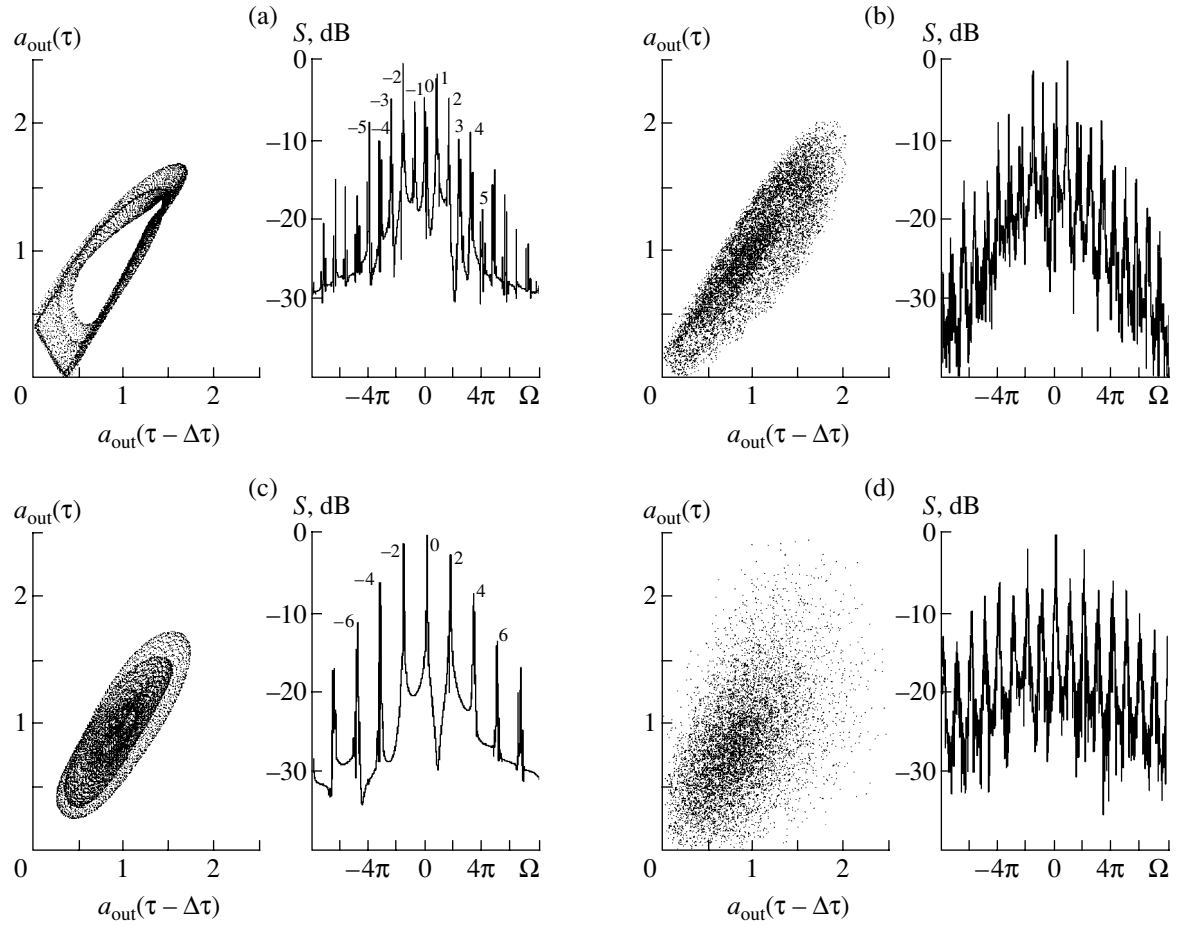


Fig. 4. Phase portraits and spectra illustrating different types of behavior observed after transition to chaos via intermittency: $\rho = 0.7$; $\delta = 1.0$; $\psi = 0.1\pi$; $\alpha = 3.9$ (a), 4.3 (b), 5.0 (c), and 5.2 (d).

to the single-soliton generation depicted in Fig. 1a is observed. Further increase in pump amplitude again leads to a transition to chaos via breakdown of quasiperiodic dynamics. Thus, a sequence of alternating regular and chaotic regimes is observed, in agreement with results obtained for different distributed systems characterized by complex dynamics [15–18, 22, 23]. The bifurcation sequence ultimately leads to a "fully developed" chaotic regime, whose phase portraits do not contain any identifiable large-scale pattern (see Fig. 4d). However, even the spectra of such regimes contain distinct components corresponding to the eigenmodes of the original system.

Overall, the complex dynamics at the center of a generation zone described above is observed for other values of the feedback parameter. The only exceptions are the cases of $\rho \ll 1$ and $\rho \sim 1$, where somewhat simpler behavior is observed ("islands" of regular dynamics in the chaotic sea disappear). However, several transitions can still be identified between chaotic attractors that have developed from various modes. The corresponding spectra differ by the locations and amplitudes of discrete peaks standing out against a noisy background.

It should also be noted that the overall dynamics becomes increasingly complex as δ increases, i.e., the frequency difference between modes decreases. Transient processes tend to take longer times, whereas the intervals of regular behavior become shorter. The regimes discussed above are increasingly difficult to identify, obviously because of competition between many closely spaced modes.

4. TRANSITION TO CHAOS NEAR THE BOUNDARIES OF A GENERATION ZONE

Finally, we discuss the complex dynamics observed near the boundaries of a generation zone, where stable (but not necessarily single-frequency) generation regimes developing from two neighboring eigenmodes coexist. The particular regime that survives after the transition is determined by the initial conditions. This behavior is typically exhibited by delayed-feedback oscillators [15–18]. The key question concerns the regime that develops from two coexisting attractors as supercriticality increases: do they merge into a single

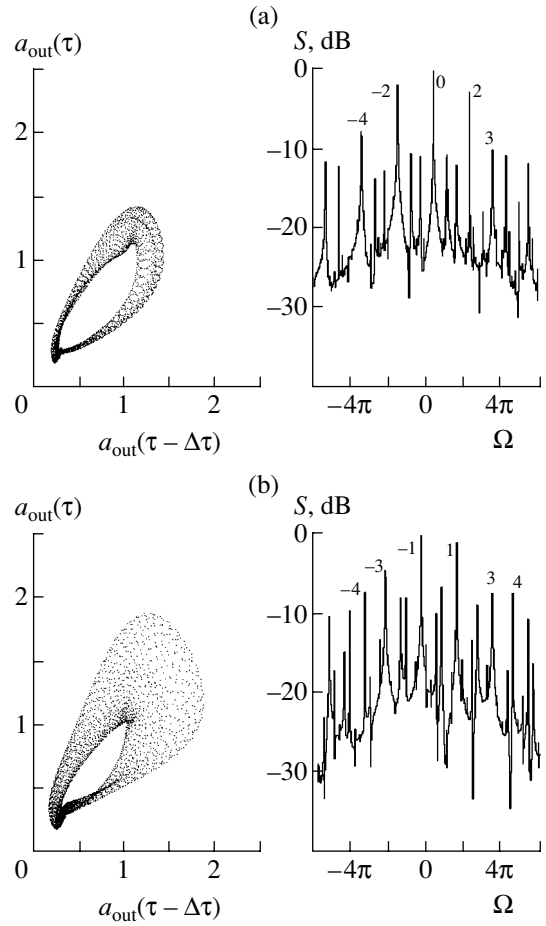


Fig. 5. Phase portraits and spectra of output signal for the bistable regime with $\rho = 0.2$, $\psi = 0.9\pi$, $\delta = 1.0$, and $\alpha = 5.37$. The coexistent self-modulation regimes develop from the modes with $n = 0$ (a) and –1 (b).

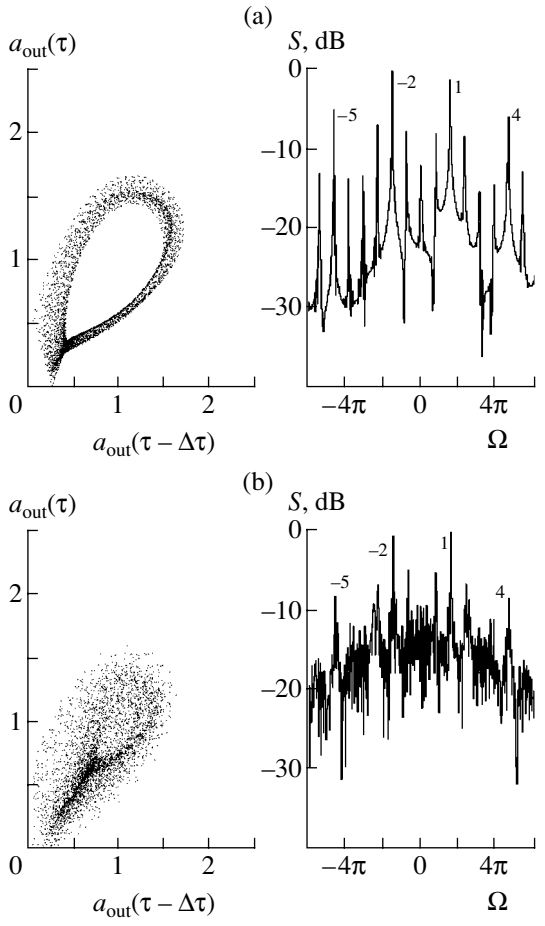


Fig. 6. Phase portraits and spectra for regimes with suppressed central modes: (a) multiple-frequency regime with $\alpha = 5.4$; (b) fully developed chaotic regime with $\alpha = 6.0$. The remaining parameters are as in Fig. 5.

attractor, or are they superseded by an essentially different regime?

To be specific, we consider the case of $\rho = 0.2$, $\delta = 1.0$, and $\psi = 0.9\pi$, when the regimes corresponding to the modes with $n = 0$ or –1 are observed, depending on initial conditions. The dynamics in question were simulated by increasing α and retaining the initial conditions. The dynamics observed under relatively weak pumping are qualitatively similar to those described above. In particular, a transition to chaos via intermittency occurs in the fundamental mode when $\alpha = 5.2$, superseded by quasiperiodic self-modulation at $\alpha = 5.28$. Figure 5 illustrates the coexistence of two quasiperiodic regimes corresponding to the same parameter values.

As α increases further, a new attractor appears, with a spectrum dominated by modes with $n = –2$ and 1 (see Fig. 6a). The spectrum, being almost symmetric about the fundamental frequency, contains distinct components at the frequencies corresponding to $n = –5$ and 4, which are almost equal to the tripled frequencies of the dominant modes. The frequencies of the remaining components are different from any eigenmode frequency. Thus, the central modes with $n = –1$ and 0 are suppressed. This regime is analogous to the mode splitting and "symmetric" cycles described in [35, 36, 39]. The output waveform is a sequence of envelope solitons having different polarities. The carrier-wave phases corresponding to neighboring solitons differ approximately by π.[2] In this respect, the regime is analogous to those discussed in Section 2 (see Fig. 1c).

The three regimes coexist within a narrow range of parameter values. Outside it, the attractors that develop from the central modes become unstable, and a single "symmetric" attractor survives. The corresponding self-modulation regime is quasiperiodic, except for a resonance observed in a narrow parameter interval due to self-mode-locking. At $\alpha \approx 5.65$, a transition to chaos occurs via breakdown of quasiperiodic dynamics. With

[2] It is clear that the phase is a function of both coordinate and time. However, it varies insignificantly within each pulse, and the difference in phase between neighboring solitons is almost equal to π.

increasing supercriticality, a transition to "fully developed" chaos analogous to that discussed above is observed (see Fig. 6b).

5. CONCLUSIONS

We have presented a detailed description of chaotic dynamics of intracavity three-wave mixing that generalized the results reported in [14, 15]. As the pump amplitude is increased, the first transition to chaos via intermittency (due to breakdown of phase locking) described in [14, 15] is followed by a complex sequence of alternating regular and chaotic self-modulation regimes in which breakdown of quasiperiodic dynamics plays a dominant role. Transitions between distinct regimes are associated with changes in spatiotemporal dynamics caused by formation of coherent structures (envelope solitons) and their interactions.

We have analyzed the complex dynamics near the boundaries of a generation zone (which was not studied in [14, 15]), where the system's behavior is further complicated by effects due to bistability, i.e., coexistence of oscillatory regimes developing from neighboring modes. With increasing supercriticality, the corresponding attractors become unstable and a new regime emerges that is symmetric about the phase-matching frequency (i.e., the gain-bandwidth center). This regime is dominated by distant side modes, whereas the central modes are suppressed. The bifurcation sequence leads to the so-called fully developed chaos, i.e., a highly irregular oscillatory regime in which no large-scale pattern can be identified in the section of the phase portrait. This behavior is characteristic of delayed-feedback distributed systems, in agreement with results reported in [15–18].

The results obtained in this study can be used to explain the complex dynamics exhibited by certain devices, particularly optical parametric oscillators. We have also observed some phenomena that occur in other systems, such as delayed-feedback traveling-wave tube oscillators [18–21], ring-cavity lasers [35–37], and "active rings" based on ferromagnetic films [40–42]. These findings suggest that the behavior demonstrated here is common to broadband delayed-feedback oscillators.

ACKNOWLEDGMENTS

This work was supported by the CRDF, project no. REC-006; the Russian Foundation for Basic Research, project no. 03-02-16192; and by the Ministry of Education of the Russian Federation, project no. A03-2.9-810.

REFERENCES

1. T. S. Akhromeeva, S. P. Kurdyumov, G. G. Malinetskiĭ, and A. A. Samarskiĭ, *Nonstationary Structures and Diffuse Chaos* (Nauka, Moscow, 1992) [in Russian].
2. M. I. Rabinovich and D. I. Trubetskov, *Introduction to the Oscillation and Wave Theory* (Nauka, Moscow, 1984; Kluwer, Dordrecht, 1989).
3. P. S. Landa, *Nonlinear Oscillations and Waves* (Kluwer, Dordrecht, 1996; Nauka, Moscow, 1997).
4. N. M. Ryskin and D. I. Trubetskov, *Nonlinear Waves* (Fizmatlit, Moscow, 2000) [in Russian].
5. M. J. Ablowitz and H. Segur, *Solitons and Inverse Spectral Transform* (SIAM, Philadelphia, 1980; Mir, Moscow, 1987).
6. D. J. Kaup, A. H. Reiman, and A. Bers, Rev. Mod. Phys. **51**, 275 (1979).
7. W. H. Louisell, *Coupled Mode and Parametric Electronics* (Wiley, New York, 1960; Inostrannaya Literatura, Moscow, 1963).
8. A. Yariv, *Quantum Electronics*, 2nd ed. (Wiley, New York, 1975; Sovetskoe Radio, Moscow, 1973).
9. V. G. Dmitriev and L. V. Tarasov, *Applied Nonlinear Optics* (Sovetskoe Radio, Moscow, 1985) [in Russian].
10. G. N. Burlak and K. Ishkabulov, Zh. Éksp. Teor. Fiz. **109**, 774 (1996) [JETP **82**, 416 (1996)].
11. *Chaotic Electronics in Telecommunications*, Ed. by M. P. Kennedy, R. Rovatti, and G. Setti (CRC Press, Boca Raton, FL, 2000).
12. Special Issue on Application of Nonlinear Dynamics to Electronic and Information Engineering, Proc. IEEE **90** (5) (2002).
13. A. S. Dmitriev and A. I. Panas, *Dynamic Chaos: New Information Carriers for Communication Systems* (Fizmatlit, Moscow, 2002) [in Russian].
14. T. V. Dmitrieva and N. M. Ryskin, Zh. Éksp. Teor. Fiz. **120**, 1517 (2001) [JETP **93**, 1314 (2001)].
15. T. V. Dmitrieva, N. M. Ryskin, and A. M. Shigaev, Nonlinear Phenom. Complex Syst. **4**, 376 (2001).
16. N. M. Ryskin and A. M. Shigaev, Zh. Tekh. Fiz. **72** (7), 1 (2002) [Tech. Phys. **47**, 795 (2002)].
17. N. M. Ryskin and V. N. Titov, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **44**, 860 (2001) [Radiophys. Quantum Electron. **44**, 793 (2001)].
18. N. M. Ryskin, V. N. Titov, S. T. Han, *et al.*, Phys. Plasmas **11**, 1194 (2004).
19. Yu. P. Bliokh, M. G. Liubarskii, V. O. Podobinskii, *et al.*, Phys. Plasmas **5**, 4061 (1998).
20. Yu. P. Bliokh, A. V. Borodkin, M. G. Lyubarskiĭ, *et al.*, Izv. Vyssh. Uchebn. Zaved., Prikl. Nelin. Din. **1** (1–2), 34 (1993).
21. Yu. P. Bliokh, M. G. Lyubarskiĭ, V. O. Podobinskiĭ, and Ya. B. Faĭnberg, Fiz. Plazmy (Moscow) **20**, 718 (1994) [Plasma Phys. Rep. **20**, 648 (1994)].
22. N. M. Ryskin, V. N. Titov, and D. I. Trubetskov, Dokl. Akad. Nauk **358**, 620 (1998) [Dokl. Phys. **43**, 90 (1998)].
23. N. M. Ryskin and V. N. Titov, Izv. Vyssh. Uchebn. Zaved., Prikl. Nelin. Din. **6** (1), 75 (1998).
24. G. L. Oppo, M. Brambilla, and L. A. Lugiato, Phys. Rev. A **49**, 2028 (1994).
25. S. Longhi, Phys. Rev. A **53**, 4488 (1996).

26. S. Longhi and A. Geraci, Phys. Rev. A **54**, 4581 (1996).
27. G. D'Alessandro, P. St. J. Russel, and A. A. Wheeler, Phys. Rev. A **55**, 3211 (1997).
28. V. J. Sanchez-Morcillo, G. J. de Valcarcel, E. Roldan, and K. Staliunas, Phys. Rev. A **56**, 3237 (1997).
29. S. Longhi, Opt. Commun. **153**, 90 (1998).
30. A. Apolinar-Irbe, N. Korneev, and J. J. Sanchez-Mondragon, Opt. Commun. **153**, 295 (1998).
31. G. J. de Valcarcel, E. Roldan, and K. Staliunas, Opt. Commun. **181**, 207 (2000).
32. P. Lodahl and M. Saffman, Opt. Commun. **184**, 493 (2000).
33. P. Suret, D. Derozier, M. Lefranc, *et al.*, Phys. Rev. A **61**, 021805 (2000).
34. S. Longhi, Opt. Commun. **204**, 339 (2002).
35. L. A. Lugiato, L. M. Narducci, E. V. Eschenazi, *et al.*, Phys. Rev. A **32**, 1563 (1985).
36. L. M. Narducci, J. R. Tredicce, L. A. Lugiato, *et al.*, Phys. Rev. A **33**, 1842 (1986).
37. E. M. Pessina, G. Bonfrate, F. Fontana, and L. A. Lugiato, Phys. Rev. A **56**, 4086 (1997).
38. N. S. Ginzburg and M. I. Petelin, Int. J. Electron. **59**, 291 (1985).
39. Ya. I. Khanin, *Principles of Laser Dynamics* (North-Holland, Amsterdam, 1995; Fizmatlit, Moscow, 1999).
40. A. A. Serga, M. P. Kostylev, B. A. Kalinikos, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **77**, 350 (2003) [JETP Lett. **77**, 300 (2003)].
41. B. A. Kalinikos, N. G. Kovshikov, M. P. Kostylev, and H. Benner, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 310 (2002) [JETP Lett. **76**, 253 (2002)].
42. S. O. Demokritov, A. A. Serga, V. E. Demidov, *et al.*, Nature **426**, 159 (2003).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 100, No. 1, 2005, pp. 218–227.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 1, 2005, pp. 242–252.

STATISTICAL, NONLINEAR, AND SOFT MATTER PHYSICS

Analytical Model of a Brownian Motor with a Fluctuating Potential

V. M. Rozenbaum and T. E. Korochkova

Institute of Surface Chemistry, National Academy of Sciences of Ukraine, Kiev, 03164 Ukraine
e-mail: vrozen@mail.kar.net

Received July 5, 2004

Abstract—We propose a model of a Brownian motor that performs a useful work against a load force F in an asymmetric periodic potential $V(x) = V(x + 2L)$ that undergoes random shifts by a half period L with a frequency γ. An arbitrarily shaped potential profile is repeated with an energy shift ΔV in both half-periods L, while the periodicity of the function $V(x)$ is ensured by its jumps at $x = 0$ and $x = L$. The boundary condition at $x = 0$ for the distribution function of a Brownian particle allows us to introduce a high and narrow potential barrier V_0 that blocks the reverse current and leads to high efficiency of the motor (the ratio of the useful work done against the load force F to the energy imparted to the particle through the potential shifts). Based on this model, we derived exact analytical expressions for the current J and the efficiency η. In the special case of piecewise-linear potentials, J and η were plotted against F and γ for various values of the parameters ΔV and V_0. We discuss the influence of the potential shape and fluctuation frequency on the main characteristics of the motor. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

At present, the nonequilibrium fluctuations in asymmetric environments that lead to a directed motion of Brownian particles even in the absence of an external field are being intensively investigated [1–3]. These issues are of methodological and practical interest in studying and constructing the so-called Brownian motors—nanomachines converting various forms of energy into mechanical energy, systems segregating nanoparticles, molecular pumps operating on the splitting energy of adenosine triphosphate, etc. In modeling such systems, an asymmetry is introduced by considering asymmetric potentials, while fluctuations are produced by time-varying deterministic or random external forces (rocking ratchets) [4–6] or by the corresponding changes in the potentials themselves (flashing ratchets) [7–9]. Considering fluctuating potentials is particularly important for biological applications [10]. Fast chemical reactions or sharply changing electric fields, which cause abrupt changes in the rate constants of chemical reactions related to directed particle transport, could be responsible for these fluctuations [11–15]. Fluctuating potentials in these models arise when the phase space of reaction coordinates can be introduced [16–18].

Asymmetric potentials that fluctuate (switch) between two states, $U^{\pm}(x)$, with a frequency γ are commonly considered. Through these fluctuations, a Brownian particle acquires a certain energy part of which dissipates when equilibrium is established in each of the potential profiles $U^{\pm}(x)$, while the other part transforms into the energy of directed motion. The main characteristics of such a Brownian motor include the current J, which determines the mean velocity of the directed motion, and the efficiency η, which characterizes the ratio of the useful work done against the external load force F to the energy spent on the potential fluctuations. A monotonic decrease in the current J with increasing load force F and a nonmonotonic behavior of the function $\eta(F)$ with a maximum η_m at a certain value of F_m are typical of a Brownian motor in a certain γ range, just as for any motor. Naturally, models with large η_m are of particular interest, especially since biological motors generally demonstrate high efficiency.

Simple potentials with a minimum number of parameters being varied, for example, sawtooth potentials characterized only by the amplitude and the asymmetry parameter, are commonly used to facilitate the computational work. Owing to the linear shape of this type of potentials, the differential equations that describe the diffusion and drift of a Brownian particle have analytical solutions, because their coefficients are constant. Astumian and Bier [7] considered the fluctuations of such potentials between two states in which only their amplitudes differed. If the potential in one of the switching states is equal to zero (see [2] for a detailed description of this case), then the particle motion in this state is purely diffusive, and, accordingly, the function $\eta(F)$ reaches a maximum at low $\eta_m \approx 0.05$ [19]. Chauwin *et al.* [9] showed that, if the potential switches between two states with identical spatial periods and if the extrema are shifted from one another in a certain way, then the Brownian motion is not involved in the generation of directed particle motion at all. In the simplest case, an asymmetric peri-

odic potential that undergoes random shifts by half a period L with a frequency γ ensures that this condition is satisfied. Clearly, one might expect an increase in η_m in this case (see model *B* of fluctuating sawtooth potentials in [20]). However, the efficiency increases significantly only in those models in which each of the potentials shifting by L contains a high barrier V_0 that blocks (under certain conditions) the reverse current [21]. An additional condition is that the flat part of the potential profile must be repeated with an energy shift ΔV in both half-periods of the potential [22]. For a periodic potential profile, these two conditions can be satisfied simultaneously if the potential is allowed to change abruptly at least at one point belonging to the interval $(0, 2L)$.

In this paper, we present the exact analytical expressions for J and η that were derived by using the above model and that were analyzed for various types of potential profiles and various parameters to elucidate the prerequisites for the high efficiency of a Brownian motor. In Section 2, we describe our model and introduce the equations that define the main characteristics of the motor. The general solution of these equations and its specific realization for high blocking barriers, for low and high potential switching frequencies, and for the special case of linear potentials are presented in Section 3. The influence of the potential shape and fluctuation frequency on the main characteristics of the motor is discussed in Section 4.

2. DESCRIPTION OF THE MODEL

The dynamics of a Brownian particle in the potentials

$$U^{\pm}(x) = V^{\pm}(x) + Fx$$

(the subscripts + and – denote the potentials with the components $V(x)$ shifted by half a period and the related quantities) is determined by two distribution functions, $\rho^{\pm}(x, t)$, that satisfy the Smoluchowski equation [23] with an additional term that describes random transitions of the particle between the potentials $U^{\pm}$ with a frequency γ:

$$\frac{\partial \rho^{\pm}(x,t)}{\partial t} = -\frac{\partial j^{\pm}(x,t)}{\partial x} - \gamma[\rho^{\pm}(x,t) - \rho^{\mp}(x,t)]. \quad (1)$$

Here, the currents $j^{\pm}(x, t)$ are defined by

$$j^{\pm}(x,t) = -D\exp[-\beta U^{\pm}(x)] \times \frac{\partial}{\partial x}\{\exp[\beta U^{\pm}(x)]\rho^{\pm}(x,t)\}, \quad (2)$$

where D is the diffusion coefficient, $\beta = (k_B T)^{-1}$ is the inverse temperature, k_B is the Boltzmann constant, and T is the absolute temperature. In a steady state, the total current $J \equiv j^{+}(x) + j^{-}(x)$ is a constant that may be represented as [22]

$$J = 2j^{+}(0) + \gamma R(L),$$
$$R(x) \equiv \int_0^x [\rho^{-}(x') - \rho^{+}(x')]dx. \quad (3)$$

Expression (3) was derived by integrating Eq. (1) over x using the conditions $\rho^{\pm}(x + L) = \rho^{\mp}(x)$ and $j^{\pm}(x + L) = j^{\mp}(x)$ that follow from the equality $V^{\pm}(x + L) = V^{\mp}(x)$. These conditions also determine the expression for the energy spent on the potential switching $U^{+} \longrightarrow U^{-} \longrightarrow U^{+}$ per unit time:

$$W_{\text{in}} = 2\gamma\int_0^L [V^{+}(x) - V^{-}(x)][\rho^{-}(x) - \rho^{+}(x)]dx. \quad (4)$$

Since the useful work done by the motor per unit time against the load force F is $W_{\text{out}} = 2FLJ$, relations (3) and (4) define the efficiency of the motor $\eta = W_{\text{out}}/W_{\text{in}}$.

In a steady state, the system of equations (1) and (2) is equivalent to fourth-order differential equations for $\rho^{\pm}(x)$ with x-dependent coefficients that are expressed in terms of the derivatives of the functions $U^{\pm}(x)$. These equations reduce to third-order equations, because the total current J is conserved as x changes. A simplification arises when piecewise-linear potentials $U^{\pm}(x)$ are used, which leads to third-order differential equations with constant coefficients. Their general solution is defined by the roots of a cubic equation, while the arbitrary constants and the sought-for current J can be determined by a numerical procedure from the continuity conditions for the functions $\rho^{\pm}(x)$ and their derivatives at the points of contact of the linear segments of the potentials [7]. The problem can be analytically considered if an arbitrarily shaped potential profile is repeated with an energy shift ΔV in both half-periods L and if the periodicity of the function $V(x)$ is ensured by its jumps at $x = 0$ and $x = L$. The differential equations for $\rho^{+}(x) \pm \rho^{-}(x)$ then become second-order equations; one of them is solved in quadratures for $U^{\pm}(x)$, while the particular solutions of the other appear explicitly in the expressions for the motor characteristics derived in the next section. In addition, this potential profile is a necessary condition for a high efficiency of the motor [22].

Let us define the function $V^{+}(x)$ in the two half-periods by the relations

$$V^{+}(x) = \begin{cases} V_0, & 0 \le x < l_0, \\ v(x), & l_0 \le x < L, \end{cases} \quad (5)$$
$$V^{+}(x + L) = v(x) - \Delta V, \quad 0 \le x < L,$$

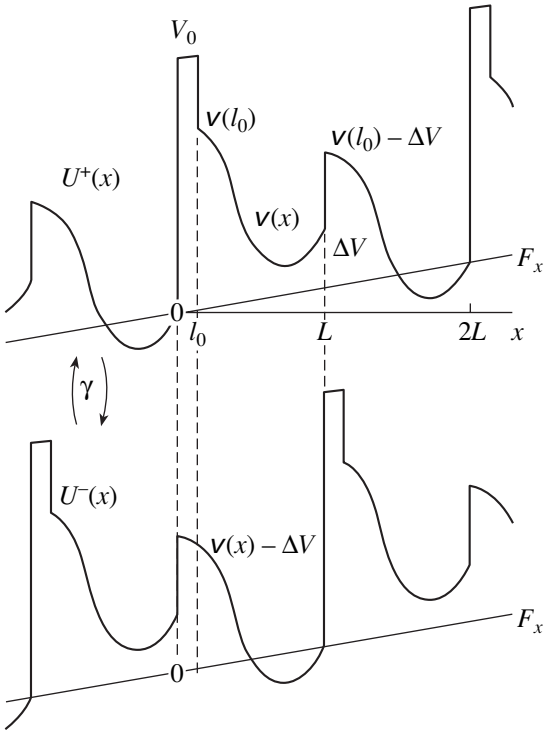


Fig. 1. The potentials $U^{\pm}(x) = V^{\pm}(x) + Fx$ that include the asymmetric periodic components $V^{\pm}(x) = V^{\pm}(x + 2L) = V^{\mp}(x + L)$ shifting by half a period L with a frequency γ and the external field of the load force F. Each potential has a high barrier V_0 in a narrow l_0 region. The flat profile $v(x)$ is repeated in both half-periods L with a shift ΔV. The periodicity of the functions $V^{\pm}(x)$ is ensured by their jumps. (The positions of the potential curves and the boundary points on them are indicated relative to the straight line Fx).

so it undergoes the jumps V_0, $V_0 - v(l_0)$, and $v(l_0) - 2\Delta V$ ($v(L) = \Delta V$, in accordance with the choice of the coordinate origin in Fig. 1) at $x = 0$, l_0, and L, respectively. In the limit $l_0 \longrightarrow 0$, the boundary conditions at $x = 0$ and $x = L$ can be easily derived from Eq. (2). These are the continuity conditions for the currents at these points and the relations that include the equality $\rho^{\pm}(x + L) = \rho^{\mp}(x)$:

$$\Lambda j^+(0) = D\{\rho^-(L) - \exp[\beta v(l_0)]\rho^+(l_0)\}, \quad (6)$$

$$\rho^+(L) = \exp\{\beta[v(l_0) - 2\Delta V]\}\rho^-(l_0), \quad (7)$$

where $\Lambda \equiv l_0 \exp(\beta V_0)$. In this limit, $V_0 l_0 \longrightarrow 0$, but arbitrary values of the parameter Λ are admissible; in particular, we may assume that $\Lambda/L \gg 1$. Therefore, relation (4) takes the form $W_{\text{in}} = 2\gamma\Delta V R(L)$, and the efficiency of the motor is given by the formula

$$\eta = \frac{FL}{\Delta V}\left[1 + \frac{2j^+(0)}{\gamma R(L)}\right]. \quad (8)$$

The efficiency η tends to unity as $FL \longrightarrow \Delta V$ if the negative current $j^+(x)$ at $x = 0$ is negligible. This condition is satisfied in the case where the barrier V_0 at $x = 0$ is high enough. Thus, the Brownian motor is highly efficient if the following two conditions [22] are satisfied simultaneously: (i) the presence of a high and narrow barrier V_0 that blocks the reverse current; and (ii) the repetition of an arbitrarily shaped flat potential profile $v(x)$ in both half-periods of the function $V^+(x)$ with an energy shift ΔV.

3. ANALYTICAL TREATMENT OF THE MAIN CHARACTERISTICS OF THE MOTOR

Let us introduce the new variables $\xi_{1,2}(x) = \rho^+(x) \mp \rho^-(x)$ that, given Eqs. (1) and (2), satisfy the following differential equations in the interval $l_0 < x < L$:

$$\{\xi_1'(x) + \beta[v'(x) + F]\xi_1(x)\}' = 2\tilde{\gamma}\xi_1(x), \quad (9)$$

$$\xi_2'(x) + \beta[v'(x) + F]\xi_2(x) = -J/D, \quad (10)$$

where $\tilde{\gamma} \equiv \gamma/D$. The general solution of Eq. (9) can be expressed in terms of two particular solutions, $\varphi_i(x)$, of the same equation,

$$\xi_1(x) = \sum_{i=1}^{2} C_i \varphi_i(x), \quad (11)$$

and contains two arbitrary constants, C_1 and C_2. Two more arbitrary constants, C_3 and J, appear in the general solution of Eq. (10):

$$\xi_2(x) = \exp\{-\beta[v(x) + Fx]\} \times \left[C_3 - \frac{J}{D}\int_{l_0}^{x} \exp\{\beta[v(x') + Fx']\}dx'\right]. \quad (12)$$

Equation (3) relates only the constants C_1 and C_2. To show this, it is convenient to introduce two functions expressed in terms of the particular solutions $\varphi_i(x)$:

$$\chi_i(x) = \varphi_i'(x) + \beta[v'(x) + F]\varphi_i(x),$$
$$\Phi_i(x) = \int_{l_0}^{x} \varphi_i(x')dx', \quad i = 1, 2. \quad (13)$$

Then, the expression for $j^+(x)$

$$2j^+(x) = J - D\sum_{i=1}^{2} C_i \chi_i(x), \quad (14)$$

follows from Eq. (2), and Eq. (3) takes the form

$$\sum_{i=1}^{2} C_i[\chi_i(0) + \tilde{\gamma}\Phi_i(L)] = 0. \quad (15)$$

Given the equality $\chi_i'(x) = 2\tilde{\gamma}\varphi_i(x)$ that follows from Eqs. (9) and (13), Eq. (15) may be rewritten in an equivalent form:

$$\sum_{i=1}^{2} C_i[\chi_i(0) + \chi_i(L)] = 0. \quad (16)$$

Thus, equalities (15) or (16) give the first equation for the constants C_1 and C_2. The second equation is determined by the normalization condition that relates the constants C_3 and J:

$$\int_{l_0}^{L} \xi_2(x)dx = C_3 Z_- - \frac{J}{D} Z_{-+} = \frac{1}{2}, \quad (17)$$

where (see Z_+ below)

$$Z_\pm \equiv \int_{l_0}^{L} \exp\{\pm\beta[V(x) + Fx]\}dx,$$

$$Z_{-+} \equiv \int_{l_0}^{L} dx \exp\{-\beta[V(x) + Fx]\} \quad (18)$$

$$\times \int_{l_0}^{x} dx' \exp\{\beta[V(x') + Fx']\}.$$

Two more equations for the four constants C_1, C_2, C_3, and J follow from the boundary conditions (6) and (7).

Having solved the system of equations (6), (7), (15) (or (16)), and (17), we may write the main characteristics of the motor as

$$\begin{aligned} j^+(0) &= \frac{D}{4\Sigma}[A - 4\tilde{\gamma}Z_+\exp(-\beta FL)\sinh(\beta\Delta V)], \\ R(L) &= \frac{1}{\Sigma}\{\Lambda[\exp(\beta\Delta V - \beta FL) - 1] + 2Z_+\exp(-\beta FL)\sinh(\beta\Delta V)\}, \\ J &= \frac{D}{2\Sigma}\{2\tilde{\gamma}\Lambda[\exp(\beta\Delta V - \beta FL) - 1] + A\}, \\ \eta &= \frac{FL}{\Delta V}\frac{\exp(\beta\Delta V - \beta FL) - 1 + (2\tilde{\gamma}\Lambda)^{-1}A}{\exp(\beta\Delta V - \beta FL) - 1 + 2\Lambda^{-1}Z_+\exp(-\beta FL)\sinh(\beta\Delta V)}, \end{aligned} \quad (19)$$

where

$$\begin{aligned} \Sigma &= 2\tilde{\gamma}\Lambda\{Z_+Z_-\exp(\beta\Delta V - \beta FL) \\ &- Z_{-+}[\exp(\beta\Delta V - \beta FL) - 1]\} \\ &+ \Lambda Z_-\{\Psi_0\exp[\beta V(l_0)] + \Psi_L\exp(2\beta\Delta V)\} \\ &+ Z_+Z_-B - Z_{-+}A, \\ A &= 2\Psi_0\exp[\beta V(l_0)] \\ &\times [\exp(-\beta FL)\cosh(\beta\Delta V) - 1] \\ &+ 2\Psi_L\exp(\beta\Delta V)[\exp(-\beta FL) - \cosh(\beta\Delta V)], \\ B &= 2\exp(-\beta FL)\{\Psi_0\exp[\beta V_0(l_0)]\cosh(\beta\Delta V) \\ &+ \Psi_L\exp(\beta\Delta V)\}, \end{aligned} \quad (20)$$

$$\Psi_j = \frac{\begin{vmatrix} \chi_1(0) + \chi_1(L)\ \varphi_1(j) \\ \chi_2(0) + \chi_2(L)\ \varphi_2(j) \end{vmatrix}}{\begin{vmatrix} \chi_1(0)\ \Phi_1(L) \\ \chi_2(0)\ \Phi_2(L) \end{vmatrix}}, \quad j = 0, L.$$

The analytical representation (19) allows us to calculate the characteristics of the motor for an arbitrarily shaped potential $V(x)$ and the parameter Λ that specifies the degree of blocking of the reverse current.

3.1. The Limit of a High Efficiency ($\Lambda/L \gg 1$)

The main result that follows from representation (19) is that, if Λ is large enough, then the efficiency of the motor tends to unity as $FL \longrightarrow \Delta V$, irrespective of the shape of the potential $V(x)$. This condition is realized when the high blocking barrier V_0 of small width l_0 is the steepest segment of the potential profile $V^+(x)$. The

efficiency in the limit $FL \longrightarrow \Delta V$ may then be approximately represented as

$$\eta \approx \frac{FL}{\Delta V}\frac{F_s - F}{F_0 - F}, \quad F_0 L = \Delta V(1+\delta_0),$$
$$F_s L = \Delta V(1-\delta_s),$$
$$\delta_0 = \frac{Z_+}{\Lambda}\frac{1-\exp(-2\beta\Delta V)}{\beta\Delta V} \ll 1, \tag{21}$$
$$\delta_s = -\frac{A}{2\tilde{\gamma}\Lambda\beta\Delta V} \ll 1,$$

and its maximum is reached for

$$\eta_m = (\sqrt{1+\delta_0} - \sqrt{\delta_0+\delta_s})^2,$$
$$F_m L = \Delta V[1+\delta_0 - \sqrt{(1+\delta_0)(\delta_0+\delta_s)}]. \tag{22}$$

The presence of a high blocking barrier allows the high efficiency of the model of a Brownian motor under consideration to be explained in terms of a strong coupling between two processes, one of which supplies energy to the motor, while the other ensures the generation of useful energy by the motor. The generalized thermodynamic driving forces of these processes are $X_2 = \beta\Delta V$ and $X_1 = -\beta FL$. The corresponding generalized currents are $J_2 = \gamma R(L)$ and $J_1 = J$, which are specified by relations (19). The entropy production in these two processes and its relationship to the efficiency of the motor may be written as [24]

$$\frac{dS}{dt} = J_1X_1 + J_2X_2 = J_2X_2(1-\eta),$$
$$\eta = -\frac{J_1X_1}{J_2X_2}. \tag{23}$$

The thermodynamic equilibrium is determined by the conditions $X_2 = X_1 = 0$ under which $J_2 = J_1 = 0$. Therefore, the generalized currents near equilibrium can be expanded in terms of small generalized forces. Assuming that these expansions are linear, we obtain

$$J_i = \sum_{j=1}^{2} L_{ij}X_j, \quad i = 1, 2, \tag{24}$$

where the expansion coefficients L_{ij} satisfy the Onsager symmetry relation $L_{12} = L_{21}$ and the inequalities $L_{11} > 0$, $L_{22} > 0$, and $L_{12}^2 \le L_{11}L_{22}$, which ensure that the quadratic form $dS/dt \ge 0$. Kedem and Caplan [24] introduced the parameter $q \equiv L_{12}/\sqrt{L_{11}L_{22}}$ $(-1 \le q \le 1)$ as a measure of coupling between the two processes in terms of which the maximum efficiency is expressed as

$$\eta_m = \frac{q^2}{(1+\sqrt{1-q^2})^2}. \tag{25}$$

When $q \longrightarrow \pm 1$, η_m tends to unity as $\eta_m \longrightarrow 1 - 2\sqrt{1-q^2}$.

Expanding relations (19) and (20) in terms of X_1 and X_2 yields

$$q = \Lambda\{[\Lambda + \tilde{\gamma}^{-1}(\Psi_0\exp[\beta v(l_0)] + \Psi_L)]$$
$$\times(\Lambda + 2Z_+)\}^{-1/2} \xrightarrow[\Lambda\to\infty]{} 1 - \frac{1}{2}(\delta_0 + \delta_s). \tag{26}$$

The small parameters δ_0 and δ_s in this expression are defined by formulas (21), which in the limit $X_{1,2} \longrightarrow 0$ take the form

$$\delta_0 = \frac{2Z_+}{\Lambda} \ll 1,$$
$$\delta_s = \frac{\Psi_0\exp[\beta v(l_0)] + \Psi_L}{\tilde{\gamma}\Lambda} \ll 1. \tag{27}$$

Thus, the presence of a high barrier blocking the reverse current ensures the coupling between the two processes whose energy is transformed by the motor near equilibrium. Relations (21) and (22) are also valid far from equilibrium where the generalized thermodynamic forces X_1 and X_2 are not small. Comparing the parameters δ_0 and δ_s defined by formulas (21) and (27), we can easily verify that they are at a minimum near equilibrium. Therefore, the highest efficiency of the motor in the presence of a blocking barrier is achieved precisely near equilibrium.

3.2. Low and High Potential Switching Frequencies

The smallness of δ_s in (21) suggests that γ is limited below. The physical meaning of this limitation is that a particle cannot pass through a high and narrow barrier V_0 over the mean lifetime γ^{-1} of the potentials $U^{\pm}$; i.e., the barrier blocks the reverse motion of the particle. The characteristic relaxation time τ in a flat potential profile $v(x) + Fx$ is given by

$$\tau = Z_+Z_-/2D \ge L^2/2D.$$

Therefore, the characteristic time in which the barrier V_0 can be overcome is estimated as τ/δ_0, and the inequality $\delta_s \ll 1$ is equivalent to the condition $\gamma^{-1} \ll \tau/\delta_0$. The behavior of solution (19) changes significantly with dimensionless parameter $\gamma\tau$. In the limiting case of low frequencies, $\gamma\tau \ll 1$, the particular solutions $\varphi_i(x)$

tend to the particular solutions of (12). This leads to the following asymptotics of the functions $\Psi_{0,L}$:

$$\begin{aligned}\Psi_0 &\xrightarrow[\gamma\to 0]{} 2Z_-^{-1}\exp[-\beta v(l_0)],\\ \Psi_L &\xrightarrow[\gamma\to 0]{} 2Z_-^{-1}\exp(-\beta\Delta V-\beta FL),\end{aligned} \tag{28}$$

which can be used to determine $\gamma R(L)$. To calculate $j^+(0)$ with the same accuracy also requires the linear (in γ) corrections to (28) that depend on the particular solutions $\varphi_i(x)$. Nevertheless, the total current calculated using Eq. (3) without applying these corrections yields a qualitatively proper behavior of the solutions for $\gamma\tau \ll 1$ considered in [22].

In the opposite limiting case of high frequencies, $\gamma\tau \geq \tilde{\gamma}L^2/2 \gg 1$, $\varphi_i(x) \approx \exp(\pm\sqrt{2\tilde{\gamma}}x)$ and the functions $\Psi_{0,L}$ become independent of the parameters of the flat potential profile:

$$\Psi_{0,L} \xrightarrow[\gamma\to\infty]{} \sqrt{2\tilde{\gamma}}. \tag{29}$$

The parameter A in (20) can then be easily expressed in terms of the slope $f_{0L} \equiv [v(l_0) - \Delta V]/L$ of the flat potential profile:

$$\begin{aligned}A &= 2\sqrt{2\tilde{\gamma}}\\ &\times\{\exp(\beta\Delta V-\beta FL)[\exp(\beta f_{0L}L)\cosh(\beta\Delta V)+1]\\ &-\exp(\beta\Delta V)[\exp(\beta f_{0L}L)+\cosh(\beta\Delta V)]\},\end{aligned} \tag{30}$$

and the maximum efficiency for $\delta_s \ll \delta_0 \ll 1$ may be written as

$$\begin{aligned}\eta_m \approx 1-2\sqrt{\delta_0}+\frac{2}{\sqrt{2\tilde{\gamma}}\Lambda}\Bigg\{&\exp(\beta f_{0L}L)\cosh(\beta\Delta V)\\ &+1-\frac{[\exp(\beta f_{0L}L)+\exp(\beta\Delta V)]\sinh(\beta\Delta V)}{\sqrt{\delta_0}\beta\Delta V}\Bigg\}.\end{aligned} \tag{31}$$

Since the parameter δ_0 defined in (21) is small, the expression in the braces is generally negative and the parameter η_m increases with $\tilde{\gamma}$, tending to the limiting value of $1-2\sqrt{\delta_0}$. The exception is the case of $f_{0L}L \sim \Delta V$ in which the expression in the braces is positive for $\Lambda/L \ll \beta\Delta V\exp(2\beta\Delta V)$. In this case, the parameter η_m shows a nonmonotonic behavior: it first reaches a maximum at some $\tilde{\gamma}$ and then decreases to $1-2\sqrt{\delta_0}$ with increasing $\tilde{\gamma}$. In the limit $\sqrt{\tilde{\gamma}}\Lambda \longrightarrow \infty$, expression (31) reduces to the asymptotics presented in [22].

3.3. Linear Potential v(x)

Solution (19) contains important information about the behavior of the main characteristics of a Brownian motor as a function of the peculiarities of the potential profile $v(x)$. To continue our analytical treatment of these characteristics, we choose the following linear function as the potential $v(x)$:

$$\begin{aligned}\beta[v(x)+Fx] &= \beta v(l_0)+f(x-l_0),\\ f &\equiv \beta F+\frac{\beta[v(L)-v(l_0)]}{L-l_0}.\end{aligned} \tag{32}$$

For $l_0 \longrightarrow 0$, the particular solutions of Eq. (9) and the functions of the parameters of $v(x)$ appearing in (19) and (20) then take the form

$$\begin{aligned}\varphi_{1,2}(x) &= \exp[(-f\pm\Delta)x/2],\quad \Delta\equiv\sqrt{f^2+8\tilde{\gamma}},\\ Z_\pm &= \pm\exp[\pm\beta v(l_0)][\exp(\pm fL)-1]/f,\\ Z_{-+} &= \frac{L}{f}-\frac{1}{f^2}[1-\exp(-fL)],\\ \Psi_{0,L} &= \frac{1}{2\sinh(L\Delta/2)}\Bigg\{\Bigg[\exp\left(\pm\frac{fL}{2}\right)\\ &+\cosh\left(\frac{L\Delta}{2}\right)\Bigg]\Delta\pm f\sinh\left(\frac{L\Delta}{2}\right)\Bigg\}.\end{aligned} \tag{33}$$

To reduce the number of parameters of the potential, let us consider two special cases: a stepwise potential with $v(l_0) = v(L)$ and a highly asymmetric sawtooth potential with $v(l_0) = 2v(L) = 2\Delta V$.

3.3.1. A stepwise potential. This special case is of interest in that the current J at $F = 0$ can be represented in a simple analytical form that characterizes the main trends in the behavior of the current as a function of several model parameters:

$$\begin{aligned}J &= \frac{D}{L^2}\frac{\tilde{\Lambda}}{\tilde{\Lambda}+[\tilde{\Lambda}+\exp(\beta\Delta V)+1]\Gamma^{-1}\coth\Gamma}\\ &\times\tanh\frac{\beta\Delta V}{2},\end{aligned} \tag{34}$$

where

$$\tilde{\Lambda}\equiv\Lambda/L,\quad \Gamma\equiv L\sqrt{\tilde{\gamma}/2}.$$

Since the potential $V(x)$ is symmetric in the absence of an additional barrier ($\Lambda = 0$) or in the absence of an energy shift between the two half-periods ($\Delta V = 0$), the current vanishes under these conditions. It also

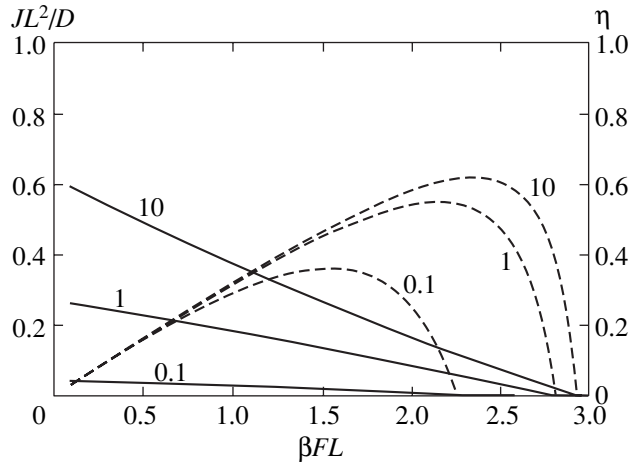


Fig. 2. Current J (solid lines) and efficiency η (dashed lines) versus load force F calculated for a stepwise potential using relations (19), (20) and (32), (33) at various $\tilde{\gamma}L^2$ (indicated near the curves) and fixed parameters $\beta\Delta V = 3$ and $\Lambda/L = 1000$.

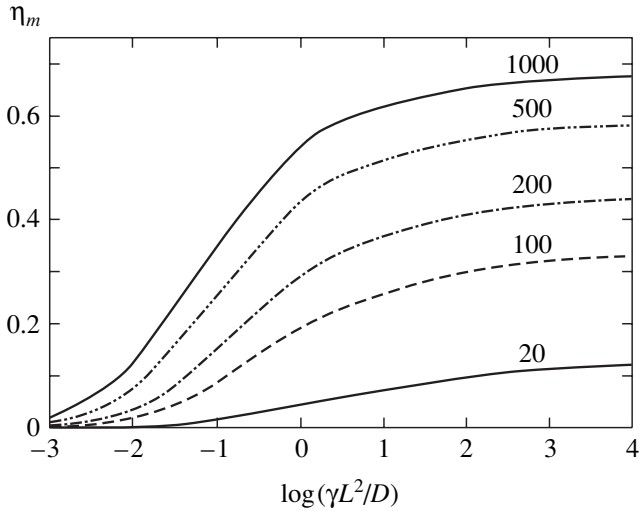


Fig. 3. $\eta_m = \max\eta(F)$ versus potential switching frequency γ for a stepwise potential at various Λ/L (indicated near the curves) and fixed parameter $\beta\Delta V = 3$.

approaches zero for $\exp(\beta\Delta V) \gg \tilde{\Lambda}$ when $\Gamma \ll 1$ or for $\exp(\beta\Delta V) \gg \Gamma\tilde{\Lambda}$ when $\Gamma \gg 1$. In the limit $\tilde{\gamma} \longrightarrow \infty$, the current approaches a nonzero value of $(D/L^2)\tanh(\beta\Delta V/2)$.

Figure 2 shows the characteristic dependences of J and η on the load force F calculated using Eqs. (19), (20) and (32), (33) at various $\tilde{\gamma}L^2$ and fixed parameters $\beta\Delta V$ and Λ/L. As γ increases, the current and the efficiency rise, with the latter showing a nonmonotonic behavior with a maximum η_m that also increases with γ. The larger the parameter Λ, the closer the monotonic dependences $\eta_m(\gamma)$ to the ideal limit $\eta = 1$ (Fig. 3).

3.3.2. A highly asymmetric sawtooth potential. In this case, the parameter ΔV simultaneously governs the asymmetry of the potential and the height of the barrier that blocks the reverse current. Therefore, the behavior of the main characteristics of the motor is nontrivial even in the absence of an additional barrier V_0 ($\Lambda = 0$). For example, in contrast to the case of a stepwise potential (see Fig. 2) or a highly asymmetric sawtooth potential with a high additional barrier (Fig. 4b), the maximum η_m of the function $\eta(F)$ changes nonmonotonically as γ increases (Fig. 4a). This dependence is clearly illustrated by Fig. 5, which shows the nonmonotonic behavior of $\eta_m(\gamma)$ in a certain range of small Λ (in agreement with the limit of large γ represented by relation (31)). The influence of an additional barrier on the efficiency of the motor is clearly seen from the dependence of the maximum of the function of two variables $\eta(F, \gamma)$ on the parameter ΔV (Fig. 6). In the range of

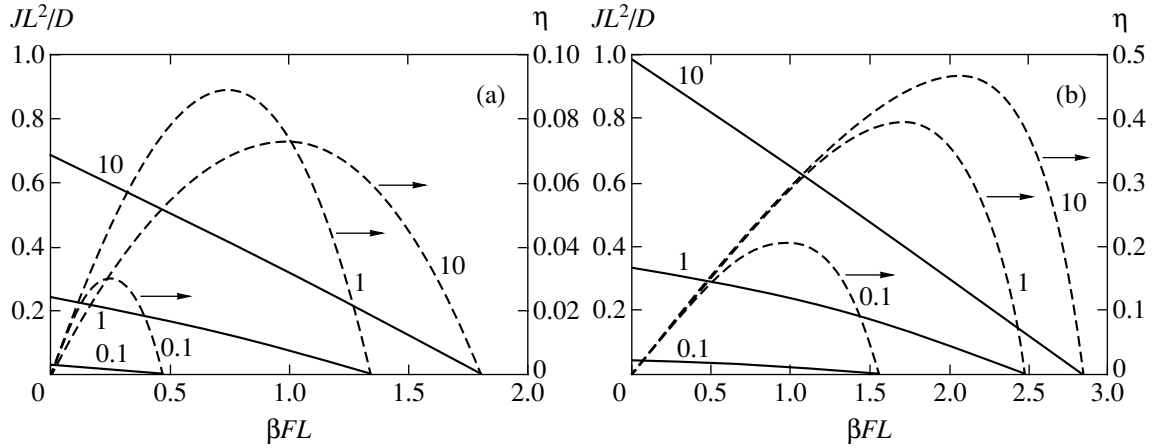


Fig. 4. Current J (solid lines) and efficiency η (dashed lines) versus load force F for a highly asymmetric sawtooth potential at various $\tilde{\gamma}L^2$ (indicated near the curves) and fixed parameters $\beta\Delta V = 3$ and $\Lambda/L = 0$ (a) and $\Lambda/L = 1000$ (b).

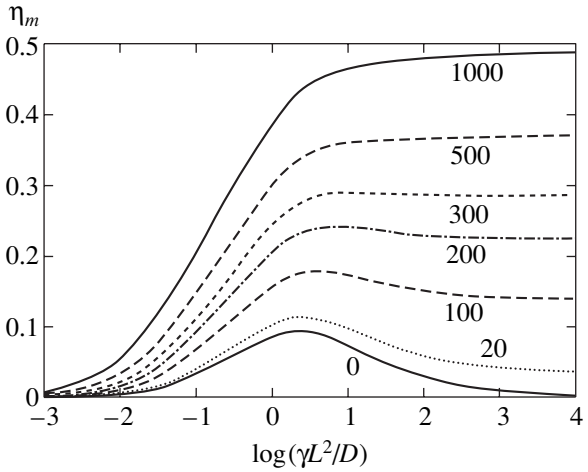


Fig. 5. $\eta_m = \max\eta(F)$ versus potential switching frequency γ for a highly asymmetric sawtooth potential at various Λ/L (indicated near the curves) and fixed parameter $\beta\Delta V = 3$.

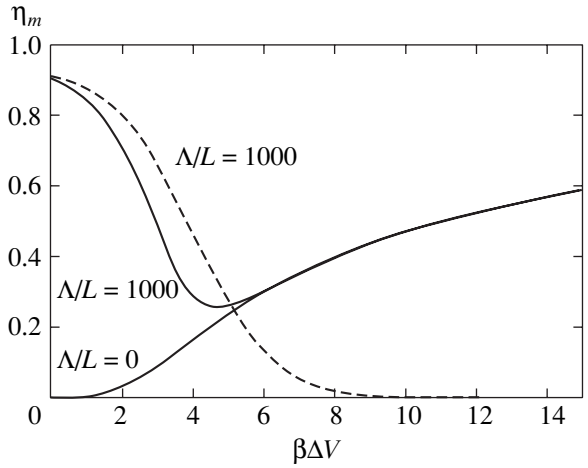


Fig. 6. $\eta_m = \max\eta(F, \gamma)$ versus potential shift in the two half-periods at various Λ/L for a highly asymmetric sawtooth potential (solid lines) and a stepwise potential (dashed line).

small ΔV, high efficiency is ensured solely by the presence of an additional barrier V_0, while the role of an additional barrier for large ΔV is insignificant compared to ΔV and the high efficiency is attributable to the blocking of the reverse current by the sawtooth potential itself. In these two limiting cases, the laws according to which η_m tends to unity are completely different. In the first case, the deviation of η_m from unity decreases exponentially rapidly with increasing V_0 (see (21) and (22)), while in the second case, it is described by the asymptotic $\eta_m \longrightarrow 1 - \ln(\beta\Delta V)/\beta\Delta V$ (the values of F_m and γ_m at which the function $\eta(F, \gamma)$ is at a maximum are determined by the asymptotics $\beta(\Delta V - F_m L) \longrightarrow \ln(\beta\Delta V)$ and $\tilde{\gamma}_m L^2 \longrightarrow \ln(\beta\Delta V)/2\beta\Delta V$). For a stepwise potential, the parameter ΔV is not related to the presence of a blocking barrier. Therefore, the efficiency η_m monotonically decreases with increasing ΔV (Fig. 6).

Expanding functions (33) in terms of the parameter $8\tilde{\gamma}/f^2 \ll 1$, we can easily represent J and η at $\Lambda = 0$ and $FL \ll \Delta V$ as

$$J = \frac{\gamma[\sinh(\beta\Delta V) - \beta\Delta V] - \tau^{-1}\sinh(\beta FL)}{2[1 + \cosh(\beta\Delta V)]},$$
$$\eta = \frac{FL}{\Delta V}\left[1 - \frac{\beta\Delta V}{\sinh(\beta\Delta V)} - \frac{1}{\gamma\tau}\frac{\sinh(\beta FL)}{\sinh(\beta\Delta V)}\right], \quad (35)$$

where

$$\tau = \left[\frac{\sinh(\beta\Delta V/2)}{\beta\Delta V/2}\right]^2 \frac{L^2}{2D}.$$

Note that expanding J in terms of small generalized forces $X_2 = \beta\Delta V$ and $X_1 = -\beta FL$ near equilibrium yields

$$J \approx \frac{1}{24}\gamma(\beta\Delta V)^3 - \frac{\beta DF}{2L}. \quad (36)$$

Since the current is proportional to the cube of ΔV, we conclude that the two processes responsible for the conversion of energy by the motor are coupled very weakly and that the maximum efficiency is very low,

$$\eta_m \longrightarrow \frac{1}{288}(\beta\Delta V)^4\tilde{\gamma}L^2.$$

Only far from equilibrium does the efficiency tend to unity as $\beta\Delta V \longrightarrow \infty$ (see Fig. 6). Expression (35) that defines η reduces to that in [22] for the case of low frequencies only if $\beta\Delta V \gg 1$ (because the current was calculated in [22] using formula (3) in which $j^+(0)$ was calculated in the zeroth approximation in γ). Introducing an additional high, narrow barrier ensures a high efficiency of a motor with a highly asymmetric sawtooth potential not only far from equilibrium, but also near it.

3.4. Continuous Potentials

An accurate analytical description of the main characteristics of a Brownian motor is possible, because the fluctuating potential undergoes abrupt changes. In addition, boundary condition (6) suggests the introduction of an additional high and narrow barrier that plays a key role in ensuring high efficiency of the motor. On the other hand, the abrupt changes in the potential may be treated as a limiting behavior of real continuous potentials. Therefore, our study would be incomplete without an analysis of the stability of the solutions obtained against the small changes in the potential

shape through which the continuous behavior becomes discontinuous.

For a Brownian particle with a diffusion coefficient D, the potential switching with a frequency γ determines the characteristic size of the region $\tilde{\gamma}^{-1/2} = \sqrt{D/\gamma}$ in which the distribution functions and the currents undergo significant changes due to the potential fluctuations. For the boundary conditions (6) and (7) to be satisfied, these distances must be much larger than the distances l at which the same quantities change significantly due to an abrupt change in the potentials. Estimating the slope of the potential profile on the segments l that shrink to a point when passing to the limit of a discontinuous potential as V_0/l, we obtain the condition

$$\gamma \ll D(\beta V_0/l)^2, \tag{37}$$

which is satisfied over a wide range of frequencies γ. Given that the currents cannot change significantly over the width l_0 of the barrier V_0 ($l_0 \ll \tilde{\gamma}^{-1/2}$), a more stringent condition arises,

$$\gamma \ll D/l_0^2. \tag{38}$$

When calculating the efficiency of a motor, we should also bear in mind that the nonzero width of the barrier V_0 makes a contribution proportional to $V_0 l_0$ to the energy W_{in} spent on the switching of the potentials $U^{\pm}$. As a result, when the potentials switch between V_0 and v_0 and when condition (38) are taken into account, the small parameter δ_0, which was represented by relation (21) in the limit $l_0 \longrightarrow 0$, can be estimated as

$$\delta_0 \sim \frac{L}{l_0}\exp(-\beta V_0) + \frac{\beta V_0 l_0}{L}\exp(-\beta v_0). \tag{39}$$

Therefore, as the barrier V_0 increases at small fixed l_0, the parameter δ_0 first decreases (approximately to $(l_0/L)\exp(-\beta v_0)$ at $\beta(V_0 - v_0) = 2\ln(L/l_0)$) and then begins to increase. Thus, the efficiency of the motor increases with Λ (see Fig. 5) only at moderately large Λ. The energy losses due to the nonzero width of the high barrier V_0 can be reduced significantly if $\exp(-\beta v_0) \ll 1$, i.e., if the transitions in the l_0 region occur between two barriers. This mechanism of the increase in the efficiency of a motor was considered in [21] for a fluctuating double-well potential.

4. DISCUSSION AND CONCLUSIONS

The main characteristics of a Brownian motor depend significantly on the shape of the fluctuating periodic potential. For a directed motion of Brownian particles to arise, this potential must introduce an asymmetry of directions, and the presence of several more factors that prevent the reverse particle motion and reduce the energy losses caused by the potential fluctuations is required for the motor to operate efficiently. Studying these factors through numerical calculations of the currents and the efficiencies for variously shaped potentials and various forms of fluctuations is very unrewarding work, because it is not only very time-consuming, but also always leaves the optimality of the results obtained open to question. We need a general approach that, on the one hand, would define the main requirements for the potential shape and, on the other hand, would admit an analytical solution of the problem and, thus, would allow an optimal mode of operation of the motor to be easily chosen. The approach proposed in this paper completely satisfies these criteria.

Indeed, the repetition of the shape of the potential $V(x)$ in both half-periods with an energy shift ΔV allows us to obtain an exact analytical solution of the problem for the fluctuations represented by random potential shifts by the same half-period by assuming that the periodicity of the function $V(x)$ is ensured by its jumps at $x = 0$ and $x = L$ and that a high and narrow barrier V_0 is introduced by the boundary condition at $x = 0$. At the same time, prerequisites for a high efficiency of the motor are created: the barrier V_0 prevents the reverse motion of a Brownian particle, while the identity of the potential shape in both half-periods at $FL \approx \Delta V$ allows the energy losses related to the relaxation after the potential switching to be avoided. Thus, only two structural elements of the potential profile are important for an optimal operation of the motor: the parameter ΔV that characterizes the potential asymmetry and is responsible for the generation of directed motion and the parameter $\Lambda = l_0\exp(\beta V_0)$ that characterizes the degree of blocking of the reverse current. The orders of magnitude of the current (for $\tilde{\gamma}L^2 \ll 1$) and the maximum possible efficiency can be estimated as follows:

$$J \sim \frac{\gamma}{2}\tanh\frac{\beta(\Delta V - FL)}{2},$$
$$\eta_m \sim 1 - \sqrt{\frac{L}{l_0}}\exp\left(-\frac{\beta V_0}{2}\right). \tag{40}$$

Here, we assume that $l_0 \longrightarrow 0$ and $V_0 \longrightarrow \infty$, but $l_0 V_0 \longrightarrow 0$ and the parameter Λ can take on arbitrary values (which must be large enough for the efficiency of the motor to be high). Clearly, these conditions reflect the idealized situation of a highly efficient motor. In reality, for barriers V_0 of nonzero width l_0, the condition for the shape identity in both half-periods is violated and energy losses deteriorating the motor characteristics arise. For the same reasons, the efficiency of the motor also decreases when abrupt changes in the potential are replaced by continuous ones. The deterioration of the characteristics can be minimized by using a double-well periodic potential considered in [21]. Indeed, if the curvature of the potential at the extreme

points is much larger than the ratio F/L, then the positions of these points may be assumed to be independent of F. Since the wells are spaced a half-period L apart, a Brownian particle makes transitions between the close neighborhoods of the minima of these wells as the potentials switch. The identity of the shapes of the remaining potential segments is no longer necessary; therefore, we can easily ensure the presence of a high barrier at $x = 0$ and satisfy the periodic boundary conditions.

The parameters ΔV and FL may be treated as the generalized forces that disturb the system from the equilibrium characterized by a zero current J. The presence of a high barrier V_0 ensures a strong coupling between the two processes whose energy are transformed by the motor. In this case, the efficiency is highest precisely near equilibrium. In contrast, a high efficiency of a Brownian motor with a highly asymmetric sawtooth potential (without an additional barrier V_0) arises from a completely different mechanism. In this case, the same parameter ΔV governs the asymmetry of the potential and the height of the barrier that blocks the reverse current. As a result, the highest possible efficiency slowly tends to unity with increasing ΔV as $\eta_m \longrightarrow 1 - \ln(\beta\Delta V)/\beta\Delta V$, i.e., reaches a maximum far from equilibrium.

REFERENCES

1. P. Reimann, Phys. Rep. **361**, 57 (2002).
2. R. D. Astumian, Science **276**, 917 (1997).
3. F. Jülicher, A. Ajdari, and J. Prost, Rev. Mod. Phys. **69**, 1269 (1997).
4. M. O. Magnasco, Phys. Rev. Lett. **71**, 1477 (1993).
5. C. R. Doering, W. Horsthemke, and J. Riordan, Phys. Rev. Lett. **72**, 2984 (1994).
6. I. M. Sokolov, Phys. Rev. E **63**, 021107 (2001).
7. R. D. Astumian and M. Bier, Phys. Rev. Lett. **72**, 1766 (1994).
8. J. Prost, J.-F. Chawin, L. Peliti, and A. Adjari, Phys. Rev. Lett. **72**, 2652 (1994).
9. J.-F. Chauwin, A. Ajdari, and J. Prost, Europhys. Lett. **27**, 421 (1994).
10. T. Y. Tsong and R. D. Astumian, Prog. Biophys. Mol. Biol. **50**, 1 (1987).
11. T. Y. Tsong and R. D. Astumian, Bioelectrochem. Bioenerg. **15**, 457 (1986).
12. V. S. Markin, T. Y. Tsong, R. D. Astumian, and R. J. Robertson, Chem. Phys. **93**, 5062 (1990).
13. V. S. Markin and T. Y. Tsong, Biophys. J. **59**, 1308 (1991).
14. V. S. Markin and T. Y. Tsong, Bioelectrochem. Bioenerg. **26**, 251 (1991).
15. Y. Chen and T. Y. Tsong, Biophys. J. **66**, 2151 (1994).
16. M. O. Magnasco, Phys. Rev. Lett. **72**, 2656 (1994).
17. R. D. Astumian and M. Bier, Biophys. J. **70**, 637 (1996).
18. D. Keller and C. Bustamante, Biophys. J. **78**, 541 (2000).
19. J. M. R. Parrondo, J. M. Blanko, F. J. Chao, and R. Brito, Europhys. Lett. **43**, 248 (1998).
20. A. Parmeggiani, F. Jülicher, A. Ajdari, and J. Prost, Phys. Rev. E **60**, 2127 (1999).
21. Yu. A. Makhnovskii, V. M. Rozenbaum, D.-Y. Yang, *et al.*, Phys. Rev. E **69**, 021102 (2004).
22. V. M. Rozenbaum, Pis'ma Zh. Éksp. Teor. Fiz. **79**, 475 (2004) [JETP Lett. **79**, 388 (2004)].
23. H. Risken, *The Fokker–Plank Equation. Methods of Solution and Applications* (Springer, Berlin, 1989).
24. O. Kedem and S. R. Caplan, Trans. Faraday Soc. **61**, 1897 (1965).

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 100, No. 1, 2005, pp. 22–30.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 1, 2005, pp. 27–36.

ATOMS, MOLECULES, OPTICS

The Above-Threshold Ionization Spectrum in a Strong Linearly Polarized Laser Field

S. P. Goreslavskii, S. V. Popruzhenko, N. I. Shvetsov-Shilovskiĭ*, and O. V. Shcherbachev

Moscow Institute of Engineering Physics, Kashirskoe sh. 31, Moscow, 115409 Russia

**e-mail: n79@narod.ru*

Received June 29, 2004

Abstract—We use the saddle-point method to derive analytical expressions for the spectral–angular probability distributions of the ionization by a strong linearly polarized laser field in Keldysh's model that are valid for an arbitrary electron energy and adiabaticity parameter. We obtain asymptotic expansions of the general formulas in the multiphoton and tunneling regimes, analyze their accuracy, and formulate the validity conditions. We provide literal estimates of the parameters that characterize the properties and evolution of the distributions.

1. INTRODUCTION

During the ionization of atoms by intense laser fields, electrons absorb more photons than the minimum number required for a transition to the continuum. This effect, called above-threshold ionization, was theoretically predicted by Keldysh in 1964 [1]. The first observation of above-threshold ionization in an experiment [2] fifteen years later gave rise to an enormous number of experimental and theoretical works. Both the above-threshold ionization proper and the accompanying generation of high-order harmonics and correlated double ionization were studied. The results of these studies were described in detail in reviews and monographs [3–6].

In the last decade, the experimental capabilities have been significantly extended in connection with the production of laser sources that generate intense femtosecond pulses with a kilohertz repetition frequency. This made it possible to accumulate reliable statistics of rare events and to study the spectral–angular probability distributions of above-threshold atomic ionization over a wide electron energy range where the recorded signal changes by 10 to 12 orders of magnitude [7]. In a linearly polarized field, the spectrum of the electrons emitted in the direction of polarization was found to have a universal shape whose basic features are determined by the field parameters and the atomic ionization potential. In the initial portion, the spectrum rapidly decreases up to energy $\varepsilon \approx (2–3)U_p$, where $U_p = F^2/4\omega^2$ is the mean vibrational electron energy in a laser field of strength F and frequency ω or the ponderomotive potential (in what follows, we use the atomic units $\hbar = m = -e = 1$). What follows next is the so-called high-energy plateau, an extended portion in which the electron yield decreases with increasing energy relatively slowly. The plateau abruptly terminates at energy $\varepsilon \approx 10U_p$. The recorded signal in the high-energy plateau region is 4 to 6 orders of magnitude lower than that in the initial portion of the spectrum.

In Keldysh's model [1, 8, 9], the ionization probability amplitude is calculated as the matrix element of the perturbation operator between a bound atomic state and a Volkov wave. Since the effect of a laser field on the motion of a free electron in Volkov states is exactly taken into account, the model describes the ionization process outside the scope of perturbation theory. On the other hand, the Volkov states do not include the interaction of the electron with the atomic core; therefore, the ionization appears as a direct transition to the continuum. The theory constructed in this way reproduces the main properties of the initial portion of the experimentally observed spectrum for above-threshold atomic ionization, but does not yield a high-energy plateau. To describe the plateau, the approach with Volkov states was modified to incorporate the interaction of the electron that happened to be in the continuum with the atomic core. In this case, the wave packet of the ionized electron calculated in the approximation of a direct transition to the continuum is taken as a zero approximation, and the scattering of the packet by the parent ion is calculated in the first order of perturbation theory in the atomic core potential [10–13]. In such calculations, the probability amplitude of the direct ionization appears in the composite rescattering matrix element.

In a zero-range potential, the direct ionization amplitude can be expressed in terms of a generalized Bessel function that is an infinite series of the products of two ordinary Bessel functions [14, 15]. The generalized Bessel functions can be easily tabulated, but they are inconvenient to perform a qualitative analysis of the spectrum for various laser-field and atomic parameters and all the more inconvenient in rescattering calculations.

1063-7761/05/10001-0022$26.00

Therefore, approximate analytical formulas that give a clear idea of the dependence of the photoionization spectra on laser–atomic parameters are of great interest. Such formulas were previously derived [1, 8, 9] by calculating the transition amplitude using the saddle-point method. Subsequently, these results have been repeatedly used and refined [16–18]. However, in all of these papers, the exponent at the saddle point was expanded in terms of the deviations of the electron energy from the maximum of the distribution. This expansion has no direct bearing on the saddle-point method and significantly limits the validity range of the results obtained. In a linearly polarized field, the spectrum is at a maximum at energy $\varepsilon = 0$. Accordingly, the distributions calculated using the energy expansion describe only the initial portion of the spectrum without covering the entire energy range with the dominant contribution of the direct ionization.

Meanwhile, the electron distributions calculated by the saddle-point method without the expansions mentioned above [19, 20] agree excellently with the calculations by other methods [12, 21] and with the experimental data on the photoionization of negative ions [22] over the entire direct ionization spectrum. As was noted above, they reproduce the main properties of the spectrum of the direct above-threshold ionization of neutral atoms by intense optical radiation. In the saddle-point method, the transition amplitude is defined by the standard formula (see (2) below) into which the explicit form of the action calculated at the complex saddle point should be substituted. In [19, 20], the emerging complex expressions were tabulated for a specific set of parameters of the problem, and the corresponding distributions were analyzed in graphical or numerical form. Below, we show that closed real expressions for the spectral–angular photoelectron distributions valid over a wide energy range can be derived for a linearly polarized field, and significant progress can be made in their analysis. Formulas of this type, but without interference and for only one direction along the field polarization, are discussed in [23].

Here, we give formulas pertaining to the ionization from a zero-range potential well. The ionization potential of the only bound *s*-state is a parameter of such a potential well.

Generalizing the results obtained to finite-range potentials in which bound states with a nonzero orbital angular momentum are possible leads to two effects. First, the common numerical factor of the spectral–angular distributions changes; naturally, this does not affect the functional dependences. Second, a known angular dependence is added. Both these factors can be taken into account just as was done previously [9, 19]. Without these complications, our formulas can be directly used to study the shape of the electron spectrum when a negative hydrogen ion is photoionized in weak or strong fields. The situation is different for neutral atoms. In this case, the approach based on Volkov states is used to describe the ionization by fairly strong low-frequency laser fields that are capable of imparting a momentum to the electron comparable to or larger than the typical atomic values. Including the long-range Coulomb field of the ion core using perturbation theory gives rise to a large factor in the photoelectron distribution for a short-range potential [24, 25]. This factor depends on the field strength, but does not depend on the electron momentum. As a result, the total ionization probability increases by several orders of magnitude, while the shape of the distribution remains unchanged.

This paper has the following structure. A general expression for the spectral–angular distributions that is valid over a wide energy range for arbitrary adiabaticity parameters is derived in Section 2. New asymptotic expansions of the distributions in the multiphoton and tunneling regimes are considered in Section 3. Our main conclusions are summarized in Section 4.

2. BASIC RELATIONS

We will consider the ionization of an *s*-state with a binding energy I in a zero-range potential by a low-frequency, $I/\omega \gg 1$, linearly polarized laser field $\mathbf{F}(t) = (F\sin\omega t, 0, 0)$. We write the rate of transition to a Volkov state with a momentum $\mathbf{p}$ due to the absorption of n photons as

$$\frac{dW_n}{d\Omega} \equiv W_n(\mathbf{p}) = 2\pi|B(\mathbf{p})|^2 p,$$

where $d\Omega$ is the element of the solid angle in the direction of the vector $\mathbf{p}$. The electron energy $\varepsilon_p = \mathbf{p}^2/2$ is determined by the energy conservation law

$$\varepsilon_p + I + U_p = n\omega = (N + k)\omega, \qquad (1)$$

where N in the second equality is the threshold number of photons required for a transition to the continuum, and $k = 0, 1, 2, \ldots$ designates the sequence of peaks in the energy spectrum. The electron energy in the $k = 0$ threshold channel is Δ, $0 < \Delta = N\omega - U_p - I < \omega$. The ionization amplitude $B(\mathbf{p})$ has the form of a single time integral over the field period [9, 19]. Since the field is a low-frequency one, we can calculate this integral by the saddle-point method. Using the latter yields

$$B(\mathbf{p}) = \frac{(2I)^{1/4}}{(2\pi)^{3/2}}\sum_{t_s}\frac{\exp(iS(\mathbf{p}, t_s))}{(S''(\mathbf{p}, t_s))^{1/2}}. \qquad (2)$$

The action $S(\mathbf{p}, t)$ is defined by

$$S(\mathbf{p}, t) = It + \frac{1}{2}\int_{-\infty}^{t} dt'\varepsilon_p(t'),$$

where $\varepsilon_p(t)$ is the time-dependent kinetic energy of the electron with the drift (canonical) momentum $\mathbf{p}$ in the

field $\mathbf{F}(t)$. The equation for the complex saddle points $\omega t_s = \omega t_0 + i\omega t_1$ is

$$\frac{1}{2}(p_x + p_F\cos\omega t_s)^2 + I + \frac{1}{2}p_\perp^2 = 0. \quad (3)$$

Here, $p_F = F/\omega = 2\sqrt{U_p}$ is the oscillation amplitude of the kinematic electron momentum (velocity) in a periodic field, and $p_\perp$ is the momentum component in a plane perpendicular to the direction of linear polarization.

The choice of the path of integration in the saddle-point method for a generalized Bessel function and the classification of the possible positions of saddle points are discussed in [14]. In our case, the path located in the band $0 < \omega t_0 < 2\pi$ of the upper half-plane ($\omega t_1 > 0$) should be drawn through two points, $\omega t_\pm$, that have the same imaginary part and symmetrically located real parts $\omega t_{0\pm} = \pi \pm d$. The presence of two terms in Eq. (2) produces interference in the spectral–angular electron distributions.

The solutions of Eq. (3) can be naturally expressed in terms of two dimensionless parameters, γ_p^2 and $\gamma_\perp^2$, similar to Keldysh's adiabaticity parameter γ^2 [1]:

$$\gamma^2 = \frac{I}{2U_p}, \quad \gamma_p^2 = \frac{I + p^2/2}{2U_p}, \quad \gamma_\perp^2 = \frac{I + p_\perp^2/2}{2U_p}.$$

Indeed, separating the real and imaginary parts of Eq. (3) yields

$$\begin{aligned}\sin(\omega t_0)\sinh(\omega t_1) &= \pm\gamma_\perp,\\ p_x + p_F\cos(\omega t_0)\cosh(\omega t_1) &= 0,\end{aligned} \quad (4)$$

where the + and – signs pertain to the points ωt_+ and ωt_-, respectively. Eliminating the trigonometric functions from Eq. (4), we find the imaginary part of the saddle points:

$$\sinh^2(\omega t_1) = \frac{\gamma_p^2 - 1}{2} + \sqrt{\left(\frac{\gamma_p^2 - 1}{2}\right)^2 + \gamma_\perp^2}. \quad (5)$$

The real part of the saddle points $\omega t_{0\pm} = \pi \pm d$ can be easily calculated with the known solution (5) using Eqs. (4):

$$d = \arccos\frac{p_x}{p_F\cosh(\omega t_1)}. \quad (6)$$

Separating the real and imaginary parts of the action $S(\mathbf{p}, t_\pm)$, performing the summation in (2) over the saddle points, and designating

$$\mathrm{Re}S_\pm = \mathrm{Re}S(\mathbf{p}, \omega t_+) - \mathrm{Re}S(\mathbf{p}, \omega t_-),$$
$$S''(\mathbf{p}, t_\pm) = \frac{4U_p\gamma}{\omega}\rho\exp(\mp\Phi),$$

we obtain the following real expression for the angular distribution of the n-photon ionization probability after simple algebraic transformations:

$$W_n(\mathbf{p}) = \frac{\omega^2 p}{2\pi^2 F\rho} \times \exp(-2\mathrm{Im}S)[1 + \cos(\mathrm{Re}S_\pm - \Phi)], \quad (7)$$

$$\mathrm{Im}S = n\omega t_1 - \frac{\varepsilon_p}{\omega}\tanh(\omega t_1)\cos^2\theta - \frac{U_p}{\omega}\sinh(\omega t_1)\cosh(\omega t_1), \quad (8)$$

$$\mathrm{Re}S_\pm = 2dn - \cos\theta\frac{pp_F}{\omega}\gamma_\perp\coth(\omega t_1) \times\left(1 + \frac{1}{2\cosh^2(\omega t_1)}\right), \quad (9)$$

$$\rho = \frac{\gamma_\perp}{\gamma}\frac{\sqrt{\sinh^4(\omega t_1) + \gamma_\perp^2}}{\sinh(\omega t_1)}, \quad (10)$$

$$\Phi = \arctan\left(\cos\theta\frac{p\tanh^2(\omega t_1)}{p_F\gamma_\perp}\right). \quad (11)$$

In Eqs. (7)–(11), the momentum p and the energy $\varepsilon_p = p^2/2$ are related to the number of absorbed photons n by the energy conservation law (1); the angle θ is measured from the direction of polarization. The n-photon ionization probability (7) with the real functions (8)–(11) is equivalent to Eq. (33) from [19] for $l = m = 0$ written in terms of the complex ionization amplitude and the complex saddle point.

The standard saddle-point method can be used to calculate the ionization amplitude if we can restrict ourselves to the quadratic expansion of the action at the saddle points, $|S'''(\omega t_s)| \ll |S''(\omega t_s)|^{3/2}$, or if the saddle points are isolated, $|S''(\omega t_s)|^{-1/2} \ll 2d$. Analysis indicates that the validity conditions in a low-frequency field, $\omega \ll I$, with a strength below the atomic one, $F \ll F_a = (2I)^{3/2}$, are satisfied for all values of the parameter γ.

The spectral–angular distributions defined by Eqs. (7)–(11) are applicable for any photoelectron energies and arbitrary adiabaticity parameters. The dependence on the electron energy, emission angle, and other parameters is present both explicitly and via the

imaginary part of the saddle point ωt_1. We see from the structure of solution (5) that ωt_1 critically depends on the sign of the combination $\gamma_p^2 - 1$. For $\gamma_p^2 > 1$, the two terms on the right-hand side of (5) are added, effectively moving the saddle point away from the real axis and reducing the probability. If, alternatively, $\gamma_p^2 < 1$, then these terms partly cancel each other, the saddle point approaches the real axis, and, hence, the probability increases. In the multiphoton regime, $\gamma_p^2 > 1$ for all energies, because $\gamma^2 > 1$. In the tunneling regime, $\gamma^2 < 1$, and the parameter γ_p^2 can be both smaller and larger than unity in different portions of the spectrum. Therefore, the behavior of the ionization probability changes near the critical energy

$$\varepsilon_{cr} = 2U_p - I, \tag{12}$$

determined by the condition $\gamma_p^2 = 1$. This conclusion is illustrated by Fig. 1, which shows how sharply the slope of the tangents changes on the opposite sides of ε_{cr}. The width of the transition region is on the order of the ionization potential.

According to the widely used semiclassical model [26], the direct ionization spectrum terminates at energy $\varepsilon_{cl} = 2U_p$, which is commonly called the classical boundary of the spectrum. In quantum theory, a sharp boundary is usually said to be absent, but the probability rapidly decreases at $\varepsilon > \varepsilon_{cl}$. We see from Fig. 1 that the rapid decrease starts from an energy ε_{cr} that is lower than the classical boundary by the ionization potential. In the deep tunneling regime, $\gamma^2 \ll 1$, the difference between the boundaries ε_{cr} and ε_{cl} is relatively small, and the transition accounts for only a small portion of the entire spectrum. However, this difference becomes important when the parameter γ^2 does not differ greatly from unity. For example, $\varepsilon_{cr} = U_p$ at $\gamma^2 = 1/2$, which is a factor of 2 lower than the classical boundary. As γ increases, the energy ε_{cr} shifts leftward and becomes equal to zero at $\gamma = 1$.

Let us also discuss some of the general properties of the spectra. Figure 2 shows the energy spectra with interference in the case where an electron is emitted at an angle to the direction of field polarization. We consider the entire range of possible emission directions (for the symmetric angles θ and $\pi - \theta$, the spectra are identical). As the emission angle θ increases, the corresponding spectra decrease increasingly fast. This trend holds not only for small emission angles [12], but also for all emission angles. Since the computational parameters for the figure correspond to tunneling ionization, $\gamma = 1/2$, the property of the spectra under discussion becomes obvious due to the presence of the factor $(1 + (\varepsilon_p \sin^2\theta)/I)^{3/2}$ in the exponent (see Eq. (26) below).

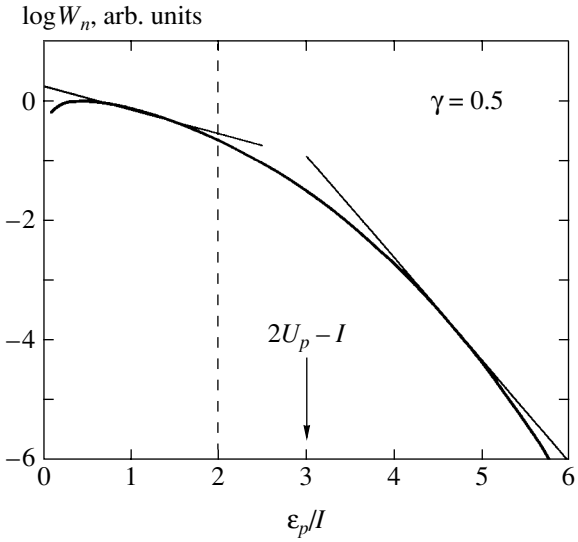


Fig. 1. The photoelectron energy spectrum in the direction of field polarization calculated using the general formulas (7)–(11) without interference. The parameters correspond to the He ionization by radiation with a frequency of $\hbar\omega = 1.58$ eV and an intensity of 0.85×10^{15} W cm^{-2}; $\gamma = 0.5$, $F/F_a = 0.06$. The position of the critical energy is indicated by an arrow.

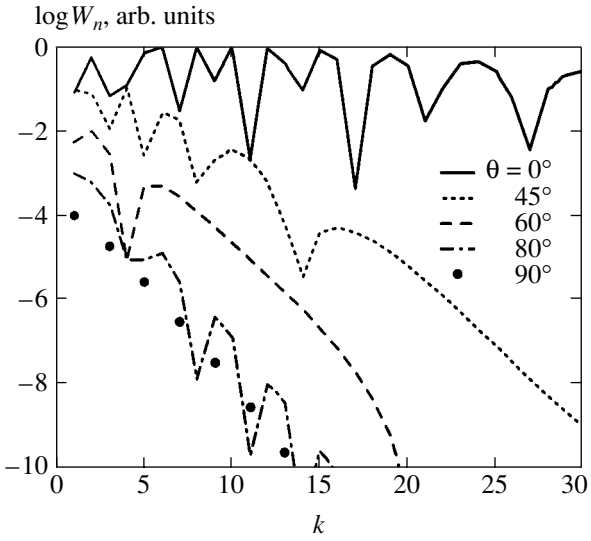


Fig. 2. The photoelectron spectra in various electron emission directions calculated with interference using the general formulas (7)–(11). The parameters are the same as those in Fig. 1.

Analysis of the interference in the spectra reveals that when an electron is emitted in the direction of polarization and at a small angle ($\theta < 45°$) to it, the separation between the neighboring minima in the energy spectrum (the "period" of the interference structure) increases with energy [12]. This property shows up most clearly at $\theta = 0$ (see Fig. 2 in [12]). Another qualitative trend can be directly seen from Fig. 2: the interference period increases with emission angle in the ini-

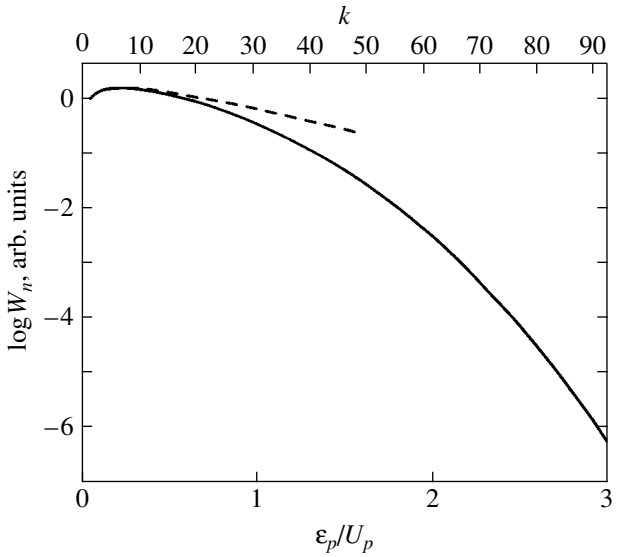


Fig. 3. The energy spectrum in the direction of polarization calculated without interference using the general formulas (7)–(11) (solid line) and formulas (14)–(17) that are applicable at low energies (dashed line). The parameters correspond to those in Fig. 1. The number of absorbed above-threshold photons is shown on the upper scale.

tial portion of the spectrum under consideration. This increase continues up to $\theta \approx 60°$ (there is only one minimum on the curve corresponding to this angle). However, as the emission angle increases further, the period begins to decrease; i.e., the number of minima in the portion of the spectrum under consideration increases. The described properties of the spectrum can also be deduced by analyzing the behavior of the derivative $d\mathrm{Re}S_{\pm}/dn$, which acts as the interference oscillation "frequency."

In the limit where the ionized electron is emitted perpendicular to the laser field direction, the interference period is constant and equal to twice the photon energy. Since the factor $[1 + \cos(\mathrm{Re}S_{\pm} - \Phi)]$ in Eq. (7) becomes equal to zero, there are no above-threshold peaks corresponding to the absorption of an odd number of photons $n = N + k$ in the spectrum [19]. This can also be seen from Eqs. (6) and (11), whence it follows that $2d = \pi$, $\Phi = 0$, and, hence, $\mathrm{Re}S_{\pm} = \pi n$ at $\theta = 90°$. For the parameters in Fig. 2, the threshold number of photons is odd ($N = 47$), and there are no above-threshold peaks with even numbers k in the spectrum in a direction perpendicular to the field. The absence of peaks in the spectrum that correspond to the absorption of an odd number of photons when a state with a zero orbital angular momentum is ionized agrees with the angular momentum and parity dipole selection rules. An above-threshold spectrum along $\theta = 90°$ with a separation of 2ω between the peaks also emerges in numerical calculations [27]. As far as we know, this property of the spectrum has not yet been observed experimentally. Concluding our discussion of the spectrum in a direction transverse to the field, we note that, in this case, $\gamma_{\perp} = \gamma_p$, $\sinh(\omega t_1) = \gamma_p$, and exponent (8) can be written as $\mathrm{Im}S = (I/\omega)g(\gamma_p)$ with the function $g(x)$ defined below by Eq. (13).

3. ASYMPTOTIC EXPANSIONS

In this section, we discuss the expansions of the exact (in terms of the saddle-point method) result (7)–(11). We derive the expansion in the range of low energies for arbitrary γ [9] and new expansions that describe the electron distribution over a wide energy range in the multiphoton, $\gamma > 1$, and tunneling, $\gamma < 1$, regimes.

3.1. Ionization to Low-Energy States

The general formulas (8)–(11) can be simplified for an arbitrary adiabaticity parameter if we restrict our analysis to the range of low energies, $\varepsilon_p < I$. Using the designation [1]

$$g(\gamma) = \left(1 + \frac{1}{2\gamma^2}\right)\mathrm{arcsinh}\gamma - \frac{\sqrt{\gamma^2+1}}{2\gamma} \tag{13}$$

and retaining only the main contributions in the expansions in terms of the small parameter $\varepsilon_p/I < 1$ in (5) and (8)–(11), we obtain

$$\mathrm{Im}S = \frac{I}{\omega}g(\gamma) + \frac{\varepsilon_p}{\omega}\left(\mathrm{arcsinh}\gamma - \frac{\gamma}{\sqrt{\gamma^2+1}}\right) + \frac{\varepsilon_p \sin^2\theta}{\omega}\frac{\gamma}{\sqrt{\gamma^2+1}}, \tag{14}$$

$$\mathrm{Re}S_{\pm} = \pi n - 4\cos\theta\sqrt{\frac{\varepsilon_p(I+2U_p)}{\omega^2}}, \tag{15}$$

$$\rho = \sqrt{\gamma^2+1}, \tag{16}$$

$$\Phi = \cos\theta\sqrt{\frac{\varepsilon_p}{I}}\frac{\gamma^2}{\gamma^2+1}. \tag{17}$$

Probability (7), whose elements can be calculated using (14)–(16), matches Eq. (53) from [9].

The spectra in the direction of polarization, $\theta = 0$, are compared in Fig. 3. The parameters correspond to the tunneling regime $I = U_p/2$, and, as we see, the agreement is excellent for $\varepsilon_p/U_p < 1/2$. As the energy increases further, the curves diverge, so difference by an order of magnitude occurs near energy $\varepsilon_p/U_p \approx 1.5$. When an electron is emitted at an angle $\theta \sim 1$ to the direction of polarization, the discrepancy between the exact formulas and the linear (in energy) expansion is observed earlier. For example, for $\varepsilon_p = U_p/2$ and $\theta = 45°$, the approximate formulas (14)–(17) underestimate the

probability by a factor of 4, while at the same energy and $\theta = 60°$, the results differ by an order of magnitude. The validity range of the low-energy expansion in the multiphoton regime also proves to be narrower (see Section 3.2).

We see from the expansion of the phase difference (15) that it is large compared to unity at low energies in both the tunneling and multiphoton regimes and changes greatly as one goes to the neighboring above-threshold peak when the electron emission direction or the intensity changes only slightly. However, the formula with the principal expansion term (15) is insufficient to accurately calculate the interference structure. The existing inaccuracy in the position of the interference minimum in the initial portion of the spectrum rapidly accumulates with increasing energy. Thus, for example, for the parameters of Fig. 3, the approximate formula (15) predicts an interference maximum at the energy that corresponds to the absorption of ten above-threshold photons, while according to the general result (9), there is a minimum here. Formula (15) can be improved by including the next expansion terms up to contributions on the order of unity. However, it is easier to calculate the interference using general formula (9), if necessary.

3.2. The Multiphoton Regime

In the multiphoton regime, $\gamma^2 > 1$, the inequality $\gamma_p^2 > 1$ holds at all electron energies. Therefore, the expansion in (5) and (8)–(11) in terms of powers of the small parameter γ_p^{-2} allows us to derive a single approximate formula for the entire above-threshold ionization spectrum. How many terms of this expansion should be included to properly calculate the contributions on the order of unity in the exponent and the phase depends on the large factors n, ε_p/ω, and U_p/ω in Eqs. (8) and (9). In the formulas given below, we restrict ourselves to the contributions that provide a relative accuracy of the exponent and the phase of $\sim\gamma_p^{-2}$:

$$\begin{aligned}\mathrm{Im}S &= n\ln\frac{2\gamma_p}{e^{1/2}} + \frac{U_p}{\omega}\left(\frac{1}{2} + \frac{3}{16\gamma_p^2}\right)\\ &- \frac{\varepsilon_p}{\omega}\left(1 - \frac{1}{2\gamma_p^2}\right) + \frac{\varepsilon_p}{\omega}\sin^2\theta\left(1 - \frac{1}{2\gamma_p^2}\right),\end{aligned} \tag{18}$$

$$\mathrm{Re}S_\pm = 2dn - 2\cos\theta\sqrt{\frac{\varepsilon_p(I + \varepsilon_p\sin^2\theta)}{\omega^2}}\left(1 + \frac{1}{\gamma_p^2}\right), \tag{19}$$

$$\begin{aligned}d &= \arccos\left(\cos\theta\sqrt{\frac{\varepsilon_p}{I + \varepsilon_p}}\right)\\ &+ \frac{\cos\theta\sqrt{\varepsilon_p(I + \varepsilon_p\sin^2\theta)}}{2\gamma_p^2(I + \varepsilon_p)},\end{aligned} \tag{20}$$

$$\rho = \frac{\gamma_\perp\gamma_p}{\gamma}, \tag{21}$$

$$\begin{aligned}\Phi &= \arctan\frac{\sqrt{\varepsilon_p}\cos\theta}{\sqrt{I + \varepsilon_p\sin^2\theta}}\\ &- \frac{\cos\theta\sqrt{\varepsilon_p(I + \varepsilon_p\sin^2\theta)}}{\gamma_p^2(I + \varepsilon_p)}.\end{aligned} \tag{22}$$

The validity ranges of the multiphoton and low-energy expansions partially overlap. The expansions in terms of the parameters $\gamma^{-2} \ll 1$ in (14)–(16) and $\varepsilon_p/I \ll 1$ in (18)–(22) yield identical results. In particular, exponent (18) takes the form

$$\begin{aligned}\mathrm{Im}S &= \frac{I}{\omega}g(\gamma) + \frac{\varepsilon_p}{\omega}\left(\ln\frac{2\gamma}{e} + \frac{3}{4\gamma^2}\right)\\ &+ \frac{\varepsilon_p}{\omega}\sin^2\theta\left(1 - \frac{1}{2\gamma^2}\right),\end{aligned} \tag{23}$$

where $g(\gamma)$ is the multiphoton limit of function (13):

$$g(\gamma) = \left(1 + \frac{1}{2\gamma^2}\right)\ln(2\gamma) - \frac{1}{2} + \frac{3}{32\gamma^2}.$$

The envelope of the spectrum calculated with the multiphoton expansion (18) virtually matches the exact result even at $\gamma = 2$: the difference is less than 10% (see Fig. 4). At $\gamma = 1$, i.e., at the boundary of the validity range, the multiphoton expansion describes the qualitative behavior of the spectrum, underestimating the ionization probability by a factor of 5 or 6, on average. We also see from Fig. 4 that the linear (in energy) approximation of the exponent in the multiphoton regime is applicable at electron energies much lower than the ionization potential. The reason is that we can substitute Eq. (18) with (23) if the discarded expansion terms are small compared to unity. The largest discarded contribution of $\sim\varepsilon_p^2/\omega I$ arises in the expansion of the logarithm in (18). It follows from the condition $\varepsilon_p^2/\omega I < 1$ that the linear (in ε_p) expansion is justified for above-threshold peaks with numbers $k < \sqrt{I/\omega}$. In a low-frequency field where $I/\omega \gg 1$, this restriction is more severe than the initially assumed validity condition for the expansion in terms of low energies $k < I/\omega$ ($I/\omega = 15$ in Fig. 4).

The term with $\sin^2\theta$ in exponent (18) describes the envelope of the angular dependence (the θ dependence is also present in the interference factor). Since $\varepsilon_p \approx k\omega$, the width of the angular distribution in the kth above-threshold peak in the multiphoton regime is estimated from (18) to be $\theta_c(k) \approx 1/\sqrt{k}$ (cf. [17]). The fairly broad

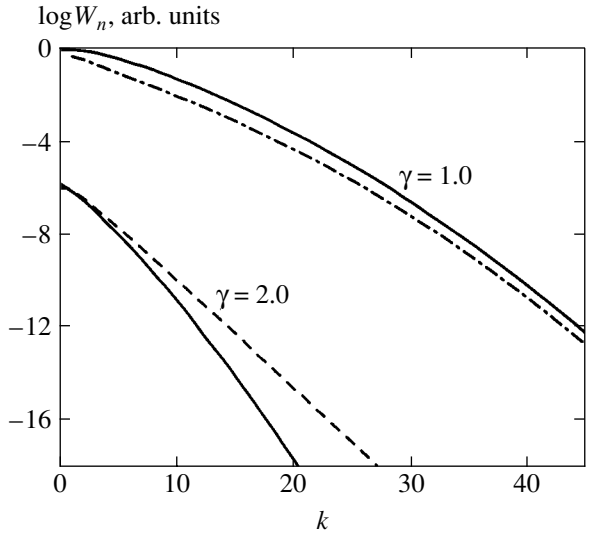


Fig. 4. The photoelectron spectra (without interference) in the direction of field polarization for the ionization of He atoms by radiation with a frequency of $\hbar\omega = 1.58$ eV at intensities of 2.0×10^{14} W cm^{-2} ($\gamma = 1.0$) and 5.3×10^{13} ($\gamma = 2.0$) W cm^{-2}. The solid and dash-dotted lines represent, respectively, the calculation using the general formulas (7)–(11) and the multiphoton expansion (18)–(22) (the dash-dotted line merges with the solid line at $\gamma = 2.0$). The dashed straight line represents the linear (in energy) approximation (14)–(17) at $\gamma = 2.0$.

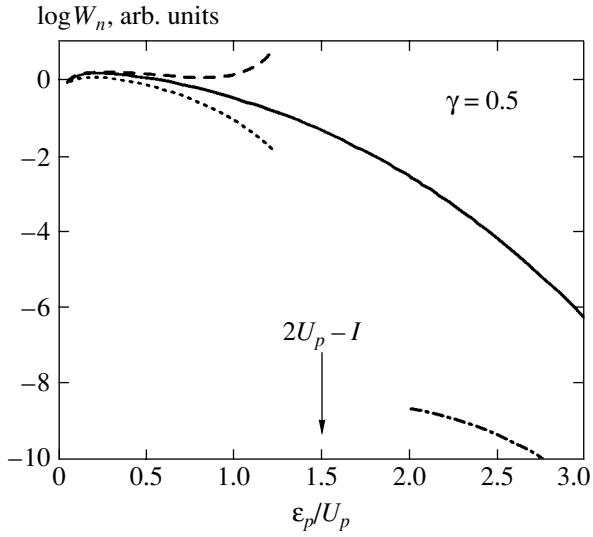


Fig. 5. The direct ionization spectrum in the direction of polarization in the tunneling regime (without the interference factor) calculated using the general and approximate formulas for ImS. The parameters are the same as those in Fig. 1. The solid, dash-dotted, dotted, and dashed lines represent the general formulas, the multiphoton expansion, the tunneling expansion (26), and the tunneling expansion (26) with the second term in the parentheses discarded.

(in the first peaks) angular distribution extends in the direction of polarization. In general, the angular distribution is at a maximum at an angle $\theta \neq 0$ due to interference [19].

Note that the condition $\gamma \gg 1$ is not sufficient to pass to the limit of perturbation theory when the probability depends on the intensity as a power law. It is necessary that the condition $U_p/\omega \ll 1$ be satisfied; this condition allows us to discard the contributions proportional to this parameter in the exponent and in the phase of the cosine. The interference structure of the distributions ceases to depend on the intensity only in this limit.

3.3. The Tunneling Regime

As was noted above, the tunneling regime is peculiar in that the parameter γ_p^2 to the left and to the right of the kinetic energy (12) take on values lower and higher than unity, respectively. Therefore, the expansions are constructed by different methods on the opposite sides of this boundary.

At fairly high electron energies, the inequality $\gamma_p^2 \gg 1$ holds and the high-energy tail of the spectral–angular distributions can be described by the multiphoton expansion formulas (18)–(22). However, it should be borne in mind that the validity range of this expansion in the tunneling regime begins at fairly high energies. Thus, for example, the condition $\gamma_p^2 \geq 3$ is satisfied for energies $\varepsilon_p \geq 5U_p$ at $\gamma^2 = 1/2$. As we see from Fig. 5, the extrapolation of the multiphoton expansion beyond the validity range toward the lower energies is invalid even for a qualitative description of the spectrum: the deviation from the real distribution reaches several orders of magnitude.

Below boundary (12), the expansion is constructed as follows. Since $\gamma_p^2 - 1 < 0$ in this portion of the spectrum, the smaller the transverse parameter $\gamma_\perp^2$ compared to unity, the stronger the cancellation of the terms on the right-hand side of Eq. (5). Therefore, it would be natural to perform an expansion in terms of the parameter $\gamma_\perp^2 \ll 1$ in Eq. (5). Under the cancellation conditions, $\sinh(\omega t_1) \approx \gamma_\perp \ll 1$, which allows the hyperbolic functions to be substituted with a power series expansion. To write the formulas derived in this section in a more compact form, it is convenient to denote $\tilde{p}_x = p_x/p_F = p\cos\theta/p_F$. Given that $1 - \gamma_p^2 = 1 - \tilde{p}_x^2 - \gamma_\perp^2$ and performing an expansion in terms of the parameter $\gamma_\perp^2 \ll 1$ in (5), we obtain

$$\omega t_1 = \frac{\gamma_\perp}{\sqrt{1-\tilde{p}_x^2}}\left\{1 - \frac{1+2\tilde{p}_x^2}{6(1-\tilde{p}_x^2)^2}\gamma_\perp^2\right\}. \qquad (24)$$

When the imaginary part of the exponent is calculated, the lowest order terms cancel each other out. Retaining

the terms on the order of $\gamma_\perp^5$, we obtain

$$\mathrm{Im}S = \frac{p_F^2}{2\omega}\left(\gamma_\perp^2\omega t_1 - \frac{1-\tilde{p}_x^2}{3}(\omega t_1)^3 - \frac{1+2\tilde{p}_x^2}{15}(\omega t_1)^5 + \ldots\right). \quad (25)$$

After the substitution of (24) into (25) and the necessary expansions in (9)–(11), we derive an approximate expression for the spectra–angular distributions in the tunneling regime:

$$\mathrm{Im}S = \frac{F_a}{3F\sqrt{1-\tilde{p}_x^2}}\left(1+\frac{p_\perp^2}{2I}\right)^{3/2} \times\left\{1-\frac{1+2\tilde{p}_x^2}{10(1-\tilde{p}_x^2)^2}\gamma_\perp^2\right\}, \quad (26)$$

$$\mathrm{Re}S_\pm = n2\arccos\left(1-\frac{\gamma_\perp^2}{2(1-\tilde{p}_x^2)}\right)\tilde{p}_x - \frac{4U_p}{\omega}\tilde{p}_x\sqrt{1-\tilde{p}_x^2}\left(\frac{3}{2}-\frac{1+2\tilde{p}_x^2}{(1-\tilde{p}_x^2)^2}\gamma_\perp^2\right), \quad (27)$$

$$\rho = \frac{\gamma_\perp}{\gamma}\sqrt{1-\tilde{p}_x^2}, \quad (28)$$

$$\tan\Phi = \tilde{p}_x\frac{\gamma_\perp}{1-\tilde{p}_x^2}. \quad (29)$$

The special cases of exponent (26) have previously been considered in the literature. Thus, setting $\tilde{p}_x = 0$ in (26), we obtain Eq. (16) from [28], while distribution (4) from [23] is derived if we set $p_\perp = 0$ in (26) and discard the second term in the braces.

There is a common validity range for the tunneling and low-energy expansions. The expansions in terms of the parameters $\gamma \ll 1$ in (14)–(17) and $\varepsilon_p/I \ll 1$ in (26)–(29) yield identical results.

The spectrum in the direction of field polarization at $\gamma = 1/2$ calculated by using the general formulas of the saddle-point method and their expansions is shown in Fig. 5. As we see from this figure, there remains a broad portion of the spectrum near the critical energy where the expansions prove to be invalid, and the calculation should be performed using the general formulas (7)–(11).

In deriving the tunneling expansion, we retained the contributions on the order of $\gamma_\perp^5$ (see (25)), which gave rise to a correction proportional to $\gamma_\perp^2$ in the parentheses of exponent (26). With this correction, the well-known [9] tunneling limit of the function $g(\gamma)$ with the quadratic contribution is obtained at $p = 0$ from (26): $g(\gamma) = (F_a/3F)(1 - \gamma^2/10)$. We see from Fig. 5 that the correction proportional to $\gamma_\perp^2$ at the very beginning of the spectrum slightly improves the agreement with the calculation based on the general formulas, but, at the same time, when the energy approaches ε_{cr}, it leads to a sharper deviation from the exact result due to the factor $(1 - \tilde{p}_x^2)^{-5/2}$. This deviation arises, because the validity condition for the tunneling expansion is violated. For $\gamma^2 \ll 1$, the inequality

$$\gamma_\perp^2 = \frac{I+\varepsilon_p\sin^2\theta}{2U_p} \ll 1$$

holds in two cases: first, for all photoelectron emission directions θ if the energy ε_p is less than or on the order of the ionization potential; and, second, for emission directions close to the direction of linear polarization when the angle θ is less than or on the order of the adiabaticity parameter γ. In the latter case, the energy ε_p can be comparable to U_p, but should not be close to the critical energy: $\varepsilon_{cr} - \varepsilon_p \gg I$. The latter restriction arises, because in deriving relation (24), $\gamma_\perp/\sqrt{1-\tilde{p}_x^2}$ is a small expansion parameter, while for $\theta = 0$ and $\varepsilon_p = \varepsilon_{cr}$, this parameter is equal to unity.

Note also that the critical energy shifts leftward and, accordingly, the validity range for the tunneling expansion narrows down as the parameter γ increases. Concurrently, the multiphoton expansion approaches the exact result (see Section 3.2).

Since the factor $\sqrt{1-\tilde{p}_x^2}$ is on the order of unity in much of the tunneling spectrum, the width of the angular distribution in the kth above-threshold peak is estimated from (26) to be

$$\theta_c(k) \approx \frac{1}{\sqrt{\gamma k}}. \quad (30)$$

Compared to the multiphoton regime, the widths $\theta_c(k)$ decrease with increasing k more slowly and, although weakly, proportional to $\sqrt{F}$, depend on the intensity. For numbers $k \approx U_p/\omega$, the width of the angular distribution becomes small, on the order of $\gamma\sqrt{F/F_a}$.

4. CONCLUSIONS

Let us summarize our main conclusions. The spectral–angular photoelectron distribution that is valid over wide ranges of energies and field strengths can be derived in explicit form by calculating the amplitude of the ionization by a strong low-frequency laser field using the saddle-point method. Our study of the analyt-

ical structure of this distribution shows that there is a critical energy under tunneling ionization conditions near which the slope of the energy spectrum changes. This energy is lower than the universally accepted classical boundary of the spectrum $2U_p$ by the ionization potential.

In the multiphoton and tunneling regimes, we obtained asymptotic expansions of the general formulas of the saddle-point method that transform to those derived previously by Perelomov *et al.* [9] in the range of low energies. In the tunneling regime, the expansions are different in form on the opposite sides of the critical energy; the spectrum is described by the multiphoton-limit formulas at higher energies. The expansions are invalid in the transition region near the critical energy, and the calculation should be performed using the general formulas. This should be taken into account when the adiabaticity parameter is moderately small compared to unity and when the transition region occupies much of the spectrum. This situation is typical of experiments on the ionization of atoms by optical radiation in the tunneling regime.

Apart from a direct analysis of the direct ionization distributions, the derived formulas can be used to analyze the accompanying processes of elastic rescattering, generation of high-order laser-frequency harmonics, and correlated double ionization. The experimental capabilities that have been extended significantly in the last decade allow these processes to be studied in elliptically polarized fields. Therefore, generalizing our results to an elliptically polarized field seems of current interest.

ACKNOWLEDGMENTS

We thank V.S. Popov for helpful discussions and N.L. Manakov, who drew our attention to [19]. This work was supported by the Russian Foundation for Basic Research (project nos. 02-02-16678 and 03-02-17112).

REFERENCES

1. L. V. Keldysh, Zh. Éksp. Teor. Fiz. **47**, 1945 (1964) [Sov. Phys. JETP **20**, 1307 (1964)].
2. P. Agostini, F. Fabre, G. Mainfray, *et al.*, Phys. Rev. Lett. **42**, 1127 (1979).
3. L. F. DiMauro and P. Agostini, Adv. At. Mol. Opt. Phys. **35**, 79 (1995).
4. M. V. Fedorov, *Atomic and Free Electrons in a Strong Light Field* (World Sci., Singapore, 1998).
5. N. B. Delone and V. P. Krainov, *Multiphoton Processes in Atoms* (Énergoatomizdat, Moscow, 1984; Springer, Berlin, 1994).
6. W. Becker, F. Grasbon, R. Kopold, *et al.*, Adv. At. Mol. Opt. Phys. **48**, 35 (2002).
7. B. Sheehy, R. Lafon, M. Widmer, *et al.*, Phys. Rev. A **58**, 3942 (1998).
8. A. I. Nikishov and V. I. Ritus, Zh. Éksp. Teor. Fiz. **50**, 255 (1966) [Sov. Phys. JETP **23**, 168 (1966)].
9. A. M. Perelomov, V. S. Popov, and M. V. Terent'ev, Zh. Éksp. Teor. Fiz. **50**, 1393 (1966) [Sov. Phys. JETP **23**, 924 (1966)].
10. M. Yu. Kuchiev, Pis'ma Zh. Éksp. Teor. Fiz. **45**, 319 (1987) [JETP Lett. **45**, 404 (1987)].
11. M. Lewenstain, K. C. Kulander, K. J. Schafer, and P. H. Bucksbaum, Phys. Rev. A **51**, 1495 (1995).
12. A. Lohr, M. Kleber, R. Kopold, and W. Becker, Phys. Rev. A **55**, R4003 (1997).
13. S. P. Goreslavskii and S. V. Popruzhenko, Zh. Éksp. Teor. Fiz. **117**, 895 (2000) [JETP **90**, 778 (2000)].
14. C. Leubner, Phys. Rev. A **23**, 2877 (1981).
15. H. Reiss, Prog. Quantum Electron. **16**, 1 (1992).
16. S. P. Goreslavskii and S. V. Popruzhenko, Zh. Éksp. Teor. Fiz. **110**, 1200 (1996) [JETP **83**, 661 (1996)].
17. V. S. Popov, Zh. Éksp. Teor. Fiz. **118**, 56 (2000) [JETP **91**, 48 (2000)].
18. V. D. Mur, S. V. Popruzhenko, and V. S. Popov, Zh. Éksp. Teor. Fiz. **119**, 893 (2001) [JETP **92**, 777 (2001)].
19. G. F. Gribakin and M. Yu. Kuchiev, Phys. Rev. A **55**, 3760 (1997).
20. G. G. Paulus, F. Zacher, H. Walther, *et al.*, Phys. Rev. Lett. **80**, 484 (1998).
21. M. V. Frolov, N. L. Manakov, E. A. Pronin, and A. F. Starace, J. Phys. B: At. Mol. Opt. Phys. **36**, L419 (2003).
22. I. Yu. Kiyan and H. Helm, Phys. Rev. Lett. **90**, 183001-1 (2003).
23. V. P. Krainov, J. Phys. B: At. Mol. Opt. Phys. **36**, L169 (2003).
24. A. I. Nikishov and V. I. Ritus, Zh. Éksp. Teor. Fiz. **52**, 223 (1967) [Sov. Phys. JETP **25**, 145 (1967)].
25. A. M. Perelomov and V. S. Popov, Zh. Éksp. Teor. Fiz. **52**, 514 (1967) [Sov. Phys. JETP **25**, 336 (1967)].
26. P. B. Corkum, N. H. Burnett, and F. Brunel, Phys. Rev. Lett. **62**, 1259 (1989).
27. E. A. Volkova, A. M. Popov, and O. V. Tikhonova, Zh. Éksp. Teor. Fiz. **120**, 1336 (2001) [JETP **93**, 1155 (2001)].
28. N. B. Delone and V. P. Kraĭnov, Usp. Fiz. Nauk **168**, 531 (1998) [Phys. Usp. **41**, 487 (1998)].

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 100, No. 1, 2005, pp. 31–38.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 1, 2005, pp. 37–44.

ATOMS, MOLECULES, OPTICS

Compton and Double-Compton Scattering Processes at Colliding Electron–Photon Beams¶

A. N. Ilyichev[a], E. A. Kuraev[b], V. V. Bytev[b], and Yu. P. Peresun'ko[c]

[a]*National Scientific and Education Center of Particle and High-Energy Physics, Belarussian State University, Minsk, 220040 Belarus*

[b]*Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia*

[c]*National Science Center, Kharkov Institute of Physics and Technology, Institute of Theoretical Physics, Kharkov, 61108 Ukraine*

e-mail: kuraev@thsun1.jinr.ru

Received July 12, 2004

Abstract—Radiative corrections to the Compton scattering cross section are calculated in the leading and next-to-leading logarithmic approximations in the case of colliding high-energy photon–electron beams. Radiative corrections to the double-Compton scattering cross section in the same experimental setup are calculated in the leading logarithmic approximation. We consider the case where no pairs are created in the final state. We show that the differential cross section can be written in the form of the Drell–Yan process cross section. Numerical values of the *K*-factor and the leading-order distribution on the scattered electron energy fraction and scattering angle are presented. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The Compton scattering process

$$\gamma(k_1) + e^-(p_1) \longrightarrow \gamma(k_2) + e^-(p_2),$$

$$k_1^2 = k_2^2 = 0, \quad p_1^2 = p_2^2 = m^2,$$

$$\begin{aligned} \kappa_1 &= 2p_1k_1 = 4\epsilon_1\omega_1, \\ \kappa_1' &= 2p_2k_1 = 2\epsilon_2\omega_1(1-c), \\ s_1 &= 2p_1p_2 = 2\epsilon_1\epsilon_2(1+c), \end{aligned} \tag{1}$$

$$\kappa_1 \sim \kappa_1' \sim s_1 \gg m^2, \quad \epsilon_2 = \frac{2\epsilon_1\omega_1}{\omega_1(1-c)+\epsilon_1(1+c)}$$

(where $\epsilon_{1,2}$, ω_1 are the energies of the initial and scattered electrons and the initial photon, $c = \cos\theta$, and θ is the angle between $\mathbf{p}_2$ and $\mathbf{k}_1$) plays an important role as a possible calibration process in high-energy photon–electron colliders [1]. Obtaining a radiation-corrected cross section of this process is the motivation of this paper. Modern methods based on the renormalization group approach in combination with the lowest-order radiative corrections (RC) makes it possible to obtain a differential cross section in the leading approximation (where $((\alpha/\pi)L)^n \sim 1$, with the "large logarithm" $L = \ln(s_1/m^2)$) and in the next-to-leading approximation (where terms on the order of $(\alpha/\pi)^n L^{n-1}$ are conserved). The accuracy of the formulas given below is therefore determined by terms on the order of

$$\frac{m^2}{\kappa_1}, \quad \frac{\alpha^2}{\pi^2}, \quad \alpha\frac{s}{M_Z^2} \tag{2}$$

compared with the terms on the order of unity and is at the per mille level for typical experimental conditions [1] $\theta \sim 1$, $\kappa_1 \sim 10$ GeV^2. Terms on the order of (2) are systematically omitted in what follows. We consider the energies of initial particles to be much less than the Z-boson mass M_Z, and, therefore, the weak corrections to the Compton effect are beyond our accuracy.

The first papers devoted to cancellation of radiative corrections to Compton scattering were published in 1952 by Brown and Feynman [2] (the contribution of virtual and soft real photon emission) and Mandl and Skyrme [3] (emission of an additional hard photon). In the work of Veltman [4], the lowest-order radiative corrections to the polarized Compton scattering were calculated in nonrelativistic kinematics. This case of kinematics was also considered in the paper by Swartz [5]. In the papers by Denner and Dittmaier [6], the lowest-order radiative corrections in the framework of the Standard Model were calculated in the case of a polarized electron and photon.

In this paper, we consider the case of high-energy electron and photon Compton scattering (with the cms energy supposed to be much higher than the electron

¶ This article was submitted by authors in English.

mass but much less than the *Z*-boson mass). We find that the cross section with radiative corrections of all orders of perturbation theory taken into account can be written in the form of the Drell–Yan process. Both leading and next-to-leading contributions are derived explicitly.

We consider the kinematics where the initial photon and electron move along the *z* axis in the opposite directions. The energy of the scattered electron is a function of its scattering angle:

$$z_0 = \frac{\varepsilon_2}{\omega_1} = \frac{2\rho}{a}, \quad a = a(c,\rho) = 1 - c + \rho(1+c), \qquad \rho = \frac{\varepsilon_1}{\omega_1}. \tag{3}$$

We now consider the kinematic case where $\rho < 1$. The case where $\rho > 1$ is considered in Appendix B.

The differential cross section in the Born approximation is given by

$$\frac{d\sigma_B}{dc}(p_1,\theta) = \frac{\pi\alpha^2 U_0}{\omega_1^2 a^2}, \quad U_0 = \frac{a}{1-c} + \frac{1-c}{a}. \tag{4}$$

In taking RC of higher orders (arising from emission of both virtual and real photons) into account, the simple relation between the scattered electron energy and the scattering angle changes, and the differential cross section in general depends on the energy fraction *z* of the scattered electron. Accepting the Drell–Yan form of the cross section, we can write it in the form

$$\frac{d\sigma}{dcdz}(p_1,\theta,z) = \int_0^1 dx D(x,L) \times \int_z^{z_0} \frac{dt}{t} D\left(\frac{z}{t},L\right) \frac{d\sigma_h}{dcdt}(xp_1,\theta,t)\left(1+\frac{\alpha}{\pi}K\right), \tag{5}$$

where the structure function $D(x, L)$ (specified below) describes the probability to find the electron (considered as a parton) inside the electron, *K* is the so-called *K*-factor, which can be calculated from the lowest RC orders, *K* is specified below (see Eqs. (8), (19), and (26)), and the "hard" cross section is

$$\frac{d\sigma_h}{dcdt}(xp_1,\theta,t) = \frac{d\sigma_B}{dc}(xp_1,\theta)\delta(t-t(x)), \tag{6}$$

$$\frac{d\sigma_B(xp_1,\theta)}{dc} = \frac{\pi\alpha^2}{\omega_1^2}\frac{1}{(1-c+\rho x(1+c))^2} \times\left(\frac{1-c}{1-c+\rho x(1+c)} + \frac{1-c+\rho x(1+c)}{1-c}\right),$$

$$t(x) = \frac{2x\rho}{1-c+\rho x(1+c)}.$$

The cross section written in the Drell–Yan form explicitly satisfies the Kinoshita–Lee–Nauenberg theorem [7]. Indeed, being integrated over the scattered electron energy fraction *z*, the structure function corresponding to the scattered electron becomes unity because

$$\int_0^1 dz \int_z^1 \frac{dt}{t} D\left(\frac{z}{t},L\right) f(t) = \int_0^1 dt f(t). \tag{7}$$

Mass singularities associated with the initial lepton structure function remain.

Therefore, our master formula for the cross section with RC taken into account is

$$\frac{d\sigma}{dzdc}(p_1,p_2) = \int_{x_0}^1 \frac{dx}{t(x)} D(x,L)\frac{d\sigma_B}{dc}(xp_1,\theta) D\left(\frac{z}{t(x)},L\right) + \frac{\alpha}{\pi}\frac{d\sigma_B(p_1,\theta)}{dc}[K_{SV}\delta(z-z_0)+K_h], \quad z = \frac{\varepsilon_2'}{\omega_1} < z_0, \quad x_0 = \frac{z(1-c)}{\rho(2-z(1+c))}, \quad L = \ln\frac{2\omega_1^2 z_0\rho(1+c)}{m^2}, \tag{8}$$

with the nonsinglet structure function *D* defined as [8]

$$D(z,L) = \delta(1-z) + \sum_{k=1}^{\infty}\frac{1}{k!}\left(\frac{\alpha L}{2\pi}\right)^k P_1(z)^{\otimes k},$$

$$P_1(z)^{\otimes k} = \underbrace{P_1 \otimes \ldots \otimes P_1(z)}_{k},$$

$$P_1 \otimes P_1(z) = \int_z^1 P_1(t)P_1\left(\frac{z}{t}\right)\frac{dt}{t}, \tag{9}$$

$$P_1(z) = \frac{1+z^2}{1-z}\theta(1-z-\Delta) + \delta(1-z)\left(2\ln\Delta + \frac{3}{2}\right), \quad \Delta \ll 1.$$

In the Conclusions (see Eq. (30)), we give the so-called "smoothed" form of the structure function.

The second term in the right-hand side of (8) collects all the nonleading contributions from the emission of virtual, soft, and hard photons, with K_{SV} given in Section 2, where the virtual and soft real contributions are considered. In Sections 3 and 4, we consider the contri-

bution from an additional hard photon emission and introduce an auxiliary parameter θ_0 to distinguish the collinear and noncollinear kinematics of photon emission. We also give the expression for the hard photon contribution K_h. The results of numerical estimation of the K-factor and leading contributions are given in Section 5. In Appendix A, we demonstrate the explicit cancellation of the θ_0 dependence. In Appendix B, we consider the kinematic case $\varepsilon_1 > \omega_1$.

2. CONTRIBUTION OF VIRTUAL AND SOFT REAL PHOTONS

To obtain the explicit form of the K-factor, we reproduce the lowest-order RC. It consists of the virtual photon emission contribution and the contribution from the real (soft and hard) photon emission. The contribution of virtual and soft photon emission was first calculated in the famous paper by Brown and Feynman [2]. The result is

$$\frac{d\sigma_{\text{virt}}}{d\sigma_B} = -\frac{\alpha}{\pi}\frac{U_1}{U_0}, \tag{10}$$

with (see [2], kinematic case II)

$$\frac{U_1}{U_0} = (1-L)\left(\frac{3}{2}+2\ln\frac{\lambda}{m}\right)+\frac{1}{2}L^2-\frac{\pi^2}{6}-K_V,$$
$$U_0 = \frac{\kappa_2}{\kappa_1}+\frac{\kappa_1}{\kappa_2}, \tag{11}$$

where K_V (the virtual photon contribution to the K-factor) is

$$K_V = -\frac{1}{U_0}$$
$$\times\left[\left(1-\frac{\kappa_2}{2\kappa_1}-\frac{\kappa_1}{\kappa_2}\right)\left(\ln^2\frac{s_1}{\kappa_1}-\ln\frac{s_1}{\kappa_1}+2\ln\frac{\kappa_2}{\kappa_1}\right)\right. \tag{12}$$
$$\left.+\left(1-\frac{\kappa_1}{2\kappa_2}-\frac{\kappa_2}{\kappa_1}\right)\left(\ln^2\frac{s_1}{\kappa_2}-\ln\frac{s_1}{\kappa_1}-\ln\frac{\kappa_1}{\kappa_2}+\pi^2\right)\right],$$

and

$$\frac{\kappa_2}{\kappa_1} = \frac{z_0(1-c)}{2\rho}, \tag{13}$$

$$\frac{s_1}{\kappa_1} = \frac{z_0(1+c)}{2}, \quad \frac{s_1}{\kappa_2} = \frac{\rho(1+c)}{1-c}. \tag{14}$$

The soft photon emission for our kinematics has the form

$$\frac{d\sigma_{\text{soft}}}{d\sigma_B} = -\frac{4\pi\alpha}{16\pi^3}$$
$$\times\int\frac{d^3k}{\omega}\left(\frac{p_1}{p_1k}-\frac{p_2}{p_2k}\right)^2_{\omega=\sqrt{\mathbf{k}^2+\lambda^2}<\Delta\epsilon\ll\epsilon_1\sim\epsilon_2}. \tag{15}$$

Standard calculations lead to the result

$$\frac{d\sigma_{\text{soft}}}{d\sigma_B} = \frac{\alpha}{\pi}\left((L-1)\ln\frac{m^2\Delta\varepsilon^2}{\lambda^2\varepsilon_1\varepsilon_2}+\frac{1}{2}L^2\right.$$
$$\left.-\frac{1}{2}\ln^2\frac{\varepsilon_1}{\varepsilon_2}-\frac{\pi^2}{3}+\mathrm{Li}_2\frac{1-c}{2}\right). \tag{16}$$

The resulting contribution to the cross section from virtual and soft real photons is independent of the fictitious "photon mass" λ and the L^2-type terms. It can be written as

$$\left(\frac{d\sigma}{dzdc}\right)_{sv} = \frac{d\sigma_{\text{virt}}+d\sigma_{\text{soft}}}{dc}\delta(z-z_0) = \frac{\alpha}{2\pi}\frac{d\sigma_B(p_1,\theta)}{dc}$$
$$\times[(L-1)(P_{1\Delta}+P_{2\Delta})+2K_{SV}]\delta(z-z_0), \tag{17}$$

where we introduce the notation

$$P_{1\Delta} = \frac{3}{2}+2\ln\frac{\Delta\varepsilon}{\varepsilon_1}, \quad P_{2\Delta} = \frac{3}{2}+2\ln\frac{\Delta\varepsilon}{\varepsilon_2}. \tag{18}$$

We can see that the terms proportional to the "large" logarithm L have a form conforming to the renormalization group prescription for the structure function. The contribution of nonleading terms is

$$K_{SV} = -\frac{\pi^2}{6}+\mathrm{Li}_2\frac{1-c}{2}-\frac{1}{2}\ln^2\frac{z_0}{\rho}+K_V. \tag{19}$$

3. CONTRIBUTION OF THE HARD COLLINEAR REAL PHOTON EMISSION

The dependence on the auxiliary parameter $\Delta\epsilon$ is eliminated when the emission of a real additional hard photon with 4-momentum k and the energy ω exceeding $\Delta\epsilon$ is taken into account.

It is convenient to consider the kinematics in which this additional photon moves within a narrow cone of the angular size $m/\epsilon_1 \ll \theta_0 \ll 1$ along the directions of the initial or scattered electrons. The contribution of these kinematic regions can be obtained using the "quasireal electron method" [9] instead of the general

(rather cumbersome) expression for the cross section of the double-Compton (DC) scattering process [3].

In the case where the collinear photon is emitted along the initial electron, the result is

$$\left(\frac{d\sigma}{dzdc}\right)_{\mathbf{k}\|\mathbf{p}_1} = \frac{\alpha}{2\pi}\int\limits_0^{1-\Delta\epsilon/\epsilon_1} dx \frac{d\sigma_B}{dc}(xp_1,\theta)$$

$$\times\left[\frac{1+x^2}{1-x}(L_1-1)+1-x\right]\delta(z-t(x)), \qquad (20)$$

$$L_1 = \ln\frac{\theta_0^2\epsilon_1^2}{m^2} = L + \ln\frac{\theta_0^2\rho}{2z_0(1+c)}.$$

When the photon is emitted along the scattered electron, we have

$$\left(\frac{d\sigma}{dzdc}\right)_{\mathbf{k}\|\mathbf{p}_2} = \frac{\alpha}{2\pi}\frac{d\sigma_B}{dc}(p_1,\theta)\int\limits_{z(1+\Delta\epsilon/\epsilon_2)}^{z_0}\frac{dt}{t}\delta(t-z_0)$$

$$\times\left[\frac{1+z^2/t^2}{1-z/t}(L_2-1)+1-\frac{z}{t}\right], \qquad (21)$$

$$L_2 = \ln\frac{\epsilon_2'^2\theta_0^2}{m^2} = L + \ln\frac{\theta_0^2 z^2}{2\rho(1+c)z_0},$$

where $z = \epsilon_2'/\omega_1 < z_0$ is the energy fraction of the scattered electron (after emission of the collinear photon).

It is convenient to write the contribution of the collinear kinematics in the form

$$\left(\frac{d\sigma_h}{dzdc}\right)_{\text{coll}} = \frac{\alpha}{2\pi}(L-1)\left[\int\limits_0^1 dx\frac{1+x^2}{1-x}\right.$$

$$\times\theta(1-x-\Delta_1)\frac{d\sigma_B(xp_1,\theta)}{dc}\delta(z-t(x))$$

$$\left.+\int\limits_z^{z_0}\frac{dt}{t}\frac{1+\frac{z^2}{t^2}}{1-\frac{z}{t}}\theta\left(1-\frac{z}{t}-\Delta_2\right)\frac{d\sigma_B(p_1,\theta)}{dc}\delta(t-z_0)\right] \qquad (22)$$

$$+\frac{df^{(1)}}{dzdc}+\frac{df^{(2)}}{dzdc},$$

where

$$\frac{df^{(1)}}{dzdc} = \frac{\alpha^3}{4\rho(1-c)\omega_1^2}\left(\frac{2-z(1+c)}{2}+\frac{2}{2-z(1+c)}\right)$$

$$\times\left[\frac{1+x^2}{1-x}\ln\frac{\rho\theta_0^2}{2z_0(1+c)}+1-x\right]_{x=x_0}\theta(1-x-\Delta_1),$$

$$\frac{df^{(2)}}{dzdc} = \frac{\alpha^3}{4\rho a\omega_1^2}\left(\frac{1-c}{a}+\frac{a}{1-c}\right) \qquad (23)$$

$$\times\left[\frac{1+z^2/t^2}{1-z/t}\ln\frac{z^2\theta_0^2}{2\rho(1+c)z_0}+1-\frac{z}{t}\right]_{t=z_0}\theta\left(1-\frac{z}{t}-\Delta_2\right),$$

$$\Delta_{1,2} = \frac{\Delta\epsilon}{\epsilon_{1,2}}.$$

We here use the relation

$$\delta(z-t(x)) = \frac{2x_0^2\rho}{z^2(1-c)}\delta(x-x_0).$$

Again, we can see that the terms containing the large logarithm L have the form conforming with the structure function. Our ansatz (5) is therefore confirmed.

The dependence on the auxiliary parameter θ_0 vanishes when the contribution of noncollinear kinematics of the additional hard photon emission is taken into account (see Appendix A).

4. NONCOLLINEAR KINEMATICS CONTRIBUTION. DOUBLE-COMPTON SCATTERING PROCESS

The general expression for the cross section of the DC scattering process

$$\gamma(k_1)+e^-(p_1)\longrightarrow\gamma(k_2)+\gamma(k)+e^-(p_2),$$

$$\kappa = 2kp_1, \quad \kappa' = 2kp_2, \qquad (24)$$

$$\kappa_2 = 2k_2p_1, \quad \kappa_2' = 2k_2p_2,$$

was obtained years ago by Mandl and Skyrme [3]. The expression for the cross section presented in this paper is exact but, unfortunately, too complicated. Instead, we use the expression for the differential cross section calculated (by the methods of chiral amplitudes [10]) under the assumption that all kinematic invariants are

Table 1. The value of K_h as a function of z and $\cos\theta$ (calculated for $\rho = 0.4$)

z\cosθ	−0.8	−0.6	−0.4	−0.2	0	0.2	0.4	0.6	0.8
0.1	–2.82	–2.61	–2.39	–2.19	–2.09	–1.89	–1.87	–2.06	–2.75
0.2	–2.77	–2.47	–2.17	–1.90	–1.65	–1.46	–1.39	–1.56	–2.30
0.3	–3.43	–2.98	–2.55	–2.14	–1.77	–1.47	–1.30	–1.38	–2.13
0.4	–4.96	–3.87	–3.23	–2.65	–2.13	–1.67	–1.34	–1.30	–2.02

Table 2. The value of $\tilde{K}_h$ as a function of y and $\cos\theta$ (calculated for $\eta = 0.064$)

y\cosθ	−0.8	−0.6	−0.4	−0.2	0	0.2	0.4	0.6	0.8
0.05	0.70	–1.97	–7.41	–15.54	–26.90	–42.70	–65.40	–100.64	–166.21
0.10	0.36	–3.20	–9.85	–18.38	–18.35				
0.15	0.03	–3.38	–1.34						
0.20	–0.20	0.29							
0.25	–0.25								

Table 3. The value of $\omega_1^2/\alpha^2 d\sigma/(dcdy)$ (the leading contribution, the first term in the right-hand side of master formula (8)) as a function of y and $\cos\theta$ (calculated for $\rho = 0.4$, $\omega_1 = 5$ GeV)

y\cosθ	−0.8	−0.6	−0.4	−0.2	0	0.2	0.4	0.6	0.8
0.1	0.211	0.237	0.265	0.299	0.345	0.413	0.526	0.754	1.450
0.2	0.337	0.357	0.378	0.405	0.445	0.508	0.618	0.850	1.576
0.3	0.703	0.669	0.643	0.634	0.644	0.685	0.782	1.013	1.784
0.4	3.883	2.153	1.554	1.264	1.113	1.054	1.090	1.296	2.122

large compared with the electron mass squared, $\kappa \sim \kappa' \sim \kappa_i \sim \kappa_i' \gg m^2$:

$$\frac{\varepsilon_2 d\sigma_0^{DC}}{d^3 p_2} = \frac{1}{2!}\frac{\alpha^3}{2\pi^2\kappa_1} R d\Phi,$$

$$d\Phi = \frac{d^3k_2}{\omega_2}\frac{d^3k}{\omega}\delta^4(p_1 + k_1 - p_2 - k_2 - k), \qquad (25)$$

$$R = s_1$$

$$\times \frac{\kappa\kappa'(\kappa^2 + \kappa'^2) + \kappa_1\kappa_1'(\kappa_1^2 + \kappa_1'^2) + \kappa_2\kappa_2'(\kappa_2^2 + \kappa_2'^2)}{\kappa\kappa'\kappa_1\kappa_1'\kappa_2\kappa_2'}.$$

The explicit expression for the contribution to the K-factor from hard photon emission K_h is

$$\frac{\alpha}{\pi}\frac{d\sigma_B}{dc}K_h = \frac{d\sigma_{\theta_0}^{DC}}{dzdc} + \frac{df^{(1)}}{dzdc} + \frac{df^{(2)}}{dzdc}, \qquad (26)$$

where

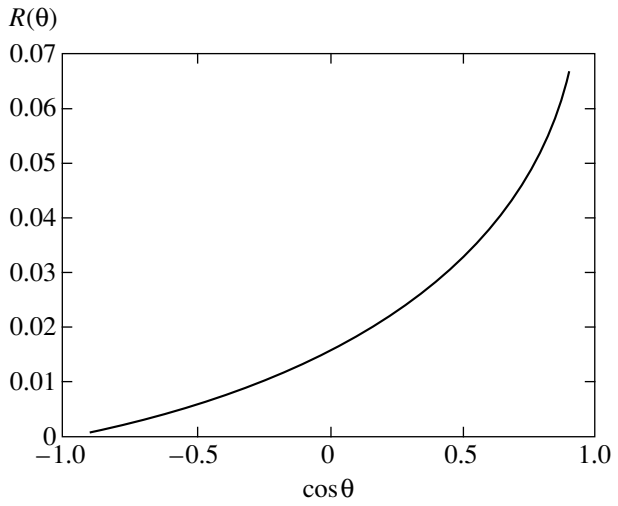
$$\frac{d\sigma_{\theta_0}^{DC}}{dzdc} = \frac{\alpha^3 z}{2!4\pi\rho}\int R d\Phi, \qquad (27)$$

and the phase volume $d\Phi$ is restricted by the conditions $\omega, \omega_2 > \Delta\varepsilon$ and the requirement that the angles between the 3-vectors $\mathbf{k}_2$, $\mathbf{k}$ and the 3-vectors $\mathbf{p}_1$, $\mathbf{p}_2$ exceed θ_0.

The values of K_h calculated numerically are given in Tables 1 and 2. We numerically and analytically find the independence of K_h of the auxiliary parameters θ_0 and $\Delta\varepsilon$ (see Appendix A).

The cross section of the DC scattering process in an inclusive experimental setup with the leading logarithmic approximation in terms of the structure functions has the form

$$d\sigma^{DC}(p_1, k_1; p_2, k, k_2) = \int_0^1 dx D(x, L) \times \int_z^1 D\left(\frac{z}{t}\right)\frac{dt}{t} d\sigma_0^{DC}\left(xp_1, k_1; \frac{tp_2}{z}, k, k_2\right), \qquad (28)$$



The leading-order radiative corrections as a $\cos\theta$ distribution (see (31)).

with the structure functions given above and

$$d\sigma_0^{DC}(p_1, k_1; p_2, k, k_2) = \frac{\alpha^3}{4\pi^2\kappa_1} \times R\frac{d^3k_2 d^3k d^3p_2}{\omega_2\omega\epsilon_2}\delta^4(p_1 + k_1 - p_2 - k_2 - k). \quad (29)$$

5. CONCLUSIONS

The characteristic form of the "reverse radiative tail" (see Tables 3 and 4) of the differential cross section vs. the energy fraction z can be reproduced if one uses the smoothed expression for nonsinglet structure functions, which includes the virtual electron pair production [11]

$$D(x, L) = \frac{\beta}{2}(1-x)^{\beta/2-1}\left(1 + \frac{3}{8}\beta\right) - \frac{\beta}{4}(1+x) + O(\beta^2), \quad \beta = \frac{2\alpha}{\pi}(L-1),$$
$$O(\beta^2) = \frac{\beta}{2}(1-x)^{\beta/2-1}\left(-\frac{1}{48}\beta^2\left(\frac{1}{3}L + \pi^2 - \frac{47}{8}\right)\right) + \frac{1}{32}\beta^2\left(-4(1+x)\ln(1-x) - \frac{1+3x^2}{1-x}\ln x - 5 - x\right). \quad (30)$$

In the figure, we give the magnitude of RC in the leading approximation

$$R(\theta) = \left(\frac{d\sigma_B}{dc}\right)^{-1}\left(\int dz\frac{d\sigma}{dz\,dc} - \frac{d\sigma_B}{dc}\right). \quad (31)$$

The results given above refer to the experimental setup without additional e^+e^-, $\mu^+\mu^-$, $\pi^+\pi^-$ real pairs in the final state.

The accuracy of the formulas given above is determined by the order of magnitude of the terms omitted (see (2)) compared to the terms on the order of unity, i.e., on the order of 0.1% for typical experimental conditions. In particular, this is why we omit the evolution effect of the K-factor terms.

The numerical value of K_h, leading contributions, and the Born cross section for different kinematic

Table 4. The value of $(\varepsilon_1^2/\alpha^2)d\tilde{\sigma}/dcdy$ (the leading contribution, the first term in the right-hand side of master formula (39)) as a function of y and $\cos\theta$ (calculated for $\omega_1 = 400$ MeV, $\varepsilon_1 = 6$ GeV)

y\cosθ	−0.8	−0.6	−0.4	−0.2	0	0.2	0.4	0.6
0.05	9.658	11.110	13.626	17.513	23.678	34.116	53.669	98.208
0.10	11.350	15.024	22.633	39.297	86.017			
0.15	13.839	23.190	56.097					
0.20	17.735	45.672						
0.25	24.303							

Table 5. Born cross section (4) (without the factor α^2/ω_1^2) for $\rho = 0.4$

cosθ	−0.8	−0.6	−0.4	−0.2	0	0.2	0.4	0.6	0.8
$\frac{\omega_1^2}{\alpha^2}\frac{d\sigma_B}{dc}$	1.779	2.038	2.365	2.796	3.389	4.266	5.721	8.669	17.881

Table 6. Born cross section (40) (without the factor α^2/ω_1^2) for $\omega_1 = 400$ MeV and $\varepsilon_1 = 6$ GeV

$\cos\theta$	−0.8	−0.6	−0.4	−0.2	0	0.2	0.4	0.6	0.8
$\frac{\omega_1^2 d\sigma_B}{\alpha^2 dc}$	93.317	60.706	49.428	44.994	44.351	47.084	54.584	72.444	129.944

regions are presented as functions of z and c in Tables 1–6.

ACKNOWLEDGMENTS

E.A.K. and V.V.B. are grateful to the Russian Foundation for Basic Research (project no. 03-02-17077) for supporting this work. All of us are grateful to S. Gevorkyan for collaboration at the beginning of this work and to S. Dittmaier for reminding us of the valuable array of earlier published papers.

APPENDIX A

Integrating the phase volume over k_2,

$$d\Phi = \frac{d^3k}{\omega}\frac{d^3k_2}{\omega_2}\delta^4(Q-k-k_2), \quad Q = p_1 + k_1 - p_2, \tag{32}$$

we can write it in the form

$$d\Phi = \frac{\omega d\omega}{\omega_1^2}\frac{2dc_1dc_2}{\sqrt{D}}\delta\Big[2\rho - \rho z(1+c) - z(1-c) - \frac{\omega}{\omega_1}(\rho(1-c_1) - z(1-c_2) + 1 + c_1)\Big], \tag{33}$$

where $D = 1 - c_1^2 - c_2^2 - c^2 - 2cc_1c_2$ and c_1, c_2 are the cosines of the respective angles between $\mathbf{k}$ and $\mathbf{p}_1$, $\mathbf{p}_2$.

For collinear kinematics, the following relations can be useful:

1. $k \approx (1-x)p_1$,

$$R_1 = R|_{\mathbf{k}\parallel\mathbf{p}_1} = \left(\frac{2x\rho}{z(1-c)} + \frac{z(1-c)}{2x\rho}\right) \times \frac{1+x^2}{(1-x)^2}\frac{1}{2\rho^2(1-c_1)x\omega_1^2},$$

$$d\Phi_1 = d\Phi|_{\mathbf{k}\parallel\mathbf{p}_1} = 2\frac{d^3k}{\omega}\delta((xp_1 + k_1 - p_2)^2) = 2\pi\frac{\rho(1-x)dxdc_1}{2-z(1+c)}\delta(x-x_0), \tag{34}$$

$$\frac{d\sigma_h^1}{dzdc} = \frac{\alpha^3 z}{2!4\pi\rho}\int R_1 d\Phi_1 = \frac{\alpha^3}{4\rho\omega_1^2(1-c)} \times \frac{1+x_0^2}{1-x_0}\left(\frac{2x_0\rho}{z(1-c)} + \frac{z(1-c)}{2x_0\rho}\right)\ln\frac{4}{\theta_0^2}.$$

In the last equation, we take into account the same contribution from the region $k \approx (1-x)p_1$.

2. In the case where $k \approx (t/z - 1)p_2$, we obtain

$$R_2 = R|_{\mathbf{k}\parallel\mathbf{p}_2} = \left(\frac{2x\rho}{z(1-c)} + \frac{z(1-c)}{2x\rho}\right) \times \frac{1+x^2}{(1-x)^2}\frac{1}{2\rho^2(1-c_1)x\omega_1^2},$$

$$d\Phi_2 = d\Phi|_{\mathbf{k}\parallel\mathbf{p}_2} = 2\frac{d^3k}{\omega}\delta((xp_1 + k_1 - p_2)^2) = 2\pi\frac{\rho(1-x)dxdc_1}{2-z(1+c)}\delta(x-x_0). \tag{35}$$

Therefore, the contribution in the case where $\mathbf{k} \parallel \mathbf{p}_2$ ($\mathbf{k}_2 \parallel \mathbf{p}_2$) has the form

$$\frac{d\sigma_h^2}{dzdc} = \frac{\alpha^3 z}{2!4\rho\omega_1^2}\int R_2 d\Phi_2 = \frac{\alpha^3}{4\rho a\omega_1^2}\left(\frac{1-c}{a} + \frac{a}{1-c}\right)\frac{1+z^2/t^2}{1-z/t}\ln\frac{4}{\theta_0^2}. \tag{36}$$

Comparing formulas (34) and (36) with (23), we can see explicit cancellation of the θ_0 dependence.

APPENDIX B

Here, we describe the different cases of kinematic regions for ρ and z.

All the above formulas were considered for $\rho < 1$, and the possible region for the variable z was determined by the inequality $x_0 < 1$,

$$z \le \frac{2\rho}{1-c+\rho(1+c)}, \tag{37}$$

which means that the lower integration limit in formula (8) is less than 1. In the case where $\rho > 1$, it is con-

Table 7. The value of y_0 and z_0 as a function of c for $\eta = 0.064$ and $\rho = 0.4$

$\cos\theta$	−0.8	−0.6	−0.4	−0.2	0	0.2	0.4	0.6	0.8
y_0	0.417	0.263	0.192	0.152	0.125	0.106	0.093	0.082	0.074
z_0	0.423	0.455	0.489	0.526	0.571	0.625	0.690	0.769	0.870

venient to introduce the new variables

$$\eta = \frac{\omega_1}{\varepsilon_1}, \quad y = \frac{\varepsilon_2'}{\varepsilon_1}, \qquad y_0 = \frac{\varepsilon_2}{\varepsilon_1} = \frac{2\eta}{1 + c + \eta(1-c)}, \quad \eta < 1. \tag{38}$$

For $\rho > 1$ (or $\eta < 1$), master equation (8) becomes

$$\frac{d\tilde{\sigma}}{dydc}(p_1, p_2) = \int_{\tilde{x}_0}^{1} \frac{dx}{\tilde{t}(x)} D(x, \tilde{L}) \frac{d\tilde{\sigma}_B(xp_1, \theta)}{dc} D\left(\frac{y}{\tilde{t}(x)}, \tilde{L}\right) + \frac{\alpha}{\pi}\frac{d\tilde{\sigma}_B(p_1, \theta)}{dc}[\tilde{K}_{SV}\delta(y - y_0) + \tilde{K}_h],$$
$$\tilde{x}_0 = \frac{y\eta(1-c)}{2\eta - y(1+c)}, \quad \tilde{L} = \ln\frac{2\varepsilon_1^2 y_0(1+c)}{m^2}, \quad \tilde{t}(x) = \frac{2\eta x}{x(1+c) + \eta(1-c)}, \tag{39}$$

with the possible values for the energy fraction y of the scattered electron given by $y \le y_0$. Born cross sections (4) and (6) and formulas for hard photon emission $\tilde{K}_{SV}$, $\tilde{K}_h$ for $\rho > 1$ follow precisely by the appropriate substitution $\rho \longrightarrow \eta^{-1}$:

$$\frac{d\tilde{\sigma}_B(xp_1, \theta)}{dc} = \frac{\pi\alpha^2}{\varepsilon_1^2}\frac{1}{(\eta(1-c) + x(1+c))^2} \times\left(\frac{\eta(1-c)}{\eta(1-c) + x(1+c)} + \frac{\eta(1-c) + x(1+c)}{\eta(1-c)}\right). \tag{40}$$

Large values of the leading contribution (see Table 4) near the kinematic bound (see Table 7) can be understood as a manifestation of the $\delta(y - y_0)$ character of the differential cross section. The y_0, z_0 dependence is given in Table 7.

REFERENCES

1. I. F. Ginzburg, G. L. Kotkin, V. G. Serbo, and V. I. Tel'nov, Pis'ma Zh. Éksp. Teor. Fiz. **34**, 514 (1981) [JETP Lett. **34**, 491 (1981)].
2. L. M. Brown and R. P. Feynman, Phys. Rev. **85**, 231 (1952); A. I. Akhiezer and V. B. Berestetski, *Quantum Electrodynamics*, 4th ed. (Nauka, Moscow, 1981; Wiley, New York, 1965), formulas (5.5.13).
3. F. Mandl and T. H. R. Skyrme, Proc. R. Soc. London, Ser. A **215**, 497 (1952).
4. H. Veltman, Phys. Rev. D **40**, 2810 (1989); **42**, 1856 (1990).
5. M. L. Swartz, Phys. Rev. D **58**, 014010 (1998); hep-ph/9711447.
6. A. Denner and S. Dittmaier, Nucl. Phys. B **407**, 43 (1993); **540**, 58 (1999); S. Dittmaier, Nucl. Phys. B **423**, 384 (1994).
7. T. Kinoshita, J. Math. Phys. **3**, 650 (1962); T. D. Lee and M. Nauenberg, Phys. Rev. B **133**, 1549 (1964).
8. S. Jadach, M. Skrzypek, and B. F. Ward, Phys. Rev. D **47**, 3733 (1993).
9. V. N. Baier, V. S. Fadin, and V. A. Khoze, Nucl. Phys. B **65**, 381 (1973).
10. F. Berends, R. Kleiss, P. de Causmaecker, *et al.*, Nucl. Phys. B **206**, 61 (1982).
11. E. A. Kuraev and V. S. Fadin, Yad. Fiz. **41**, 733 (1985) [Sov. J. Nucl. Phys. **41**, 466 (1985)].

Journal of Experimental and Theoretical Physics, Vol. 100, No. 1, 2005, pp. 39–49.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 1, 2005, pp. 45–55.

ORDER, DISORDER, AND PHASE TRANSITIONS IN CONDENSED SYSTEMS

Macroscopic Description of Superconducting Pairing under Repulsive Interaction

V. I. Belyavsky* and Yu. V. Kopaev**

Lebedev Institute of Physics, Russian Academy of Sciences, Leninskii pr. 53, Moscow, 119991 Russia
**e-mail: vib@spu.ac.ru*
***e-mail: kopaev@sci.lebedev.ru*
Received September 8, 2004

Abstract—The order parameter of a quasi-two-dimensional electron system in which superconducting pairing with large momentum and repulsive interaction occurs represents an alternating function of the momentum of the relative motion of a pair. This function vanishes at points on the Fermi contour (FC) segments inside each equivalent domain of kinematic constraint. The superfluid density is proportional to the area of such a domain. When the pairing interaction is repulsive, the order parameter can be described by a two-component complex function of coordinates that represents a solution to a system of two Ginzburg–Landau differential equations. The characteristic size of a pair, which is determined by the effective mass rather than the velocity of an electron near the FC, as well as the penetration depth and the correlation length for the fluctuations of the order parameter are obtained. The order parameter forms incommensurable quasi-periodic structures that arise under long-wavelength fluctuations of the phase and correspond to antiferromagnetically ordered current circulations. The pairing channel considered provides a qualitative description for the characteristic features of superconductivity in cuprates. © *2005 Pleiades Publishing, Inc.*

1. Superconductivity in quasi-two-dimensional (2D) cuprates, which arises under the doping of a parent insulator with antiferromagnetic (AF) ordering, exists in a limited interval of doping concentrations $x_* < x < x^*$. The temperature of superconducting (SC) transition T_C attains its maximum at a point $x = x_{opt}$ in this interval. In the underdoped regime ($x_* < x < x_{opt}$), the superfluid density of the SC condensate of singlet pairs is small, $n_s \sim (x - x_*)$, and the Fermi contour (FC) is determined by the total concentration of charge carriers $(1 - x)$ and may have several points (nodes) at which the SC gap vanishes [1, 2]. When $x_* < x < x_{opt}$, noncoherent pairs exist in the domain of a pseudogap state $T_C < T < T^*$, where the temperature T^* at which the pseudogap emerges corresponds to the association of charge carriers into pairs, whereas T_C corresponds to the emergence of phase coherence in the system of already existing pairs [3]. Since the phase stiffness is small, well-developed fluctuations arise that suppress the long-range order in a 2D system, so that T_C can be considered as the Berezinskii–Kosterlitz–Thouless transition temperature, which corresponds to the generation of uncoupled vortices and antivortices of the phase of the SC order parameter [4]. The weak interaction between electrons in adjacent copper–oxygen CuO_2 planes stabilizes the system with respect to the long-wavelength fluctuations of the order parameter, thus accounting for the dependence of T_C on the number of CuO_2 layers in the unit cell [5]. The short-range AF order admits the existence of orbital currents (current circulations) [6], which can manifest themselves as a density wave (DDW) [7] or a DDW fluctuation [8] in a system with a *d*-wave SC gap.

A quite general phenomenological approach [9, 10] to the unification of AF and SC orderings shows that, along with the Cooper SC pairing with zero momentum of a pair and a dielectric pairing with a large AF vector, there necessarily exist other pairing channels, each of which can become the dominant channel under certain conditions. One of such channels is a singlet SC pairing with large momentum of a pair. The states of such pairs belong to the class of eigenstates of the Hubbard Hamiltonian [11]. Under the assumption that the hopping of pairs between adjacent lattice sites is admitted, these states may lead to superconductivity for the positive energy of on-site correlation [12]. The SC pairing with nonzero momentum gives rise to an inhomogeneous state similar to the Fulde–Ferrel–Larkin–Ovchinnikov (FFLO) state [13, 14] with spatial modulation of the modulus and phase of the order parameter [15].

Both attraction due to electron–phonon interaction [16, 17] and repulsion [18] are considered as the pairing mechanisms. The well-known features of the physics of cuprates can be qualitatively explained in terms of the pairing with large momentum under Coulomb repulsion [19]. In the present paper, we develop a macroscopic approach to the description of superconductivity in cuprates. This approach is based on the Ginzburg–Landau equations [20] derived for this pairing channel in a way similar to that used in [21] for the Bardeen–

Cooper–Schrieffer (BCS) model [22]. Due to the fluctuation phenomena, the applicability of the mean-field approximation is limited to the overdoping domain $x_{opt} < x < x^*$. However, the results concerning the structural features of the order parameter can be extended to the low doping domain.

2. The momenta $\mathbf{k}_\pm$ of particles that compose a pair with the total momentum $\mathbf{K} = \mathbf{k}_+ + \mathbf{k}_-$ and the momentum of relative motion $\mathbf{k} = (\mathbf{k}_+ - \mathbf{k}_-)/2$ in a degenerate 2D electron system have a limited domain of variation Ξ_K. The shape and the size of this domain are determined by the Fermi contour and the momentum $\mathbf{K}$ [23]. The domain Ξ_K of kinematic constraints includes one-particle states whose linear combination gives the wavefunction of the relative motion of a pair. The boundary that separates filled and vacant parts of the domain Ξ_K with $\mathbf{K} \neq 0$ is generally given by isolated points. However, for certain $\mathbf{K}$ close to the doubled Fermi momentum in the direction of $\mathbf{K}$ and for a special shape of the FC, this boundary may be given by finite segments of the FC that satisfy the mirror nesting condition [19] and form a pair Fermi contour (PFC). This contour plays the same role for the relative motion of a pair as the FC plays for electrons and holes [19].

The pairing with total momentum $\mathbf{K} \neq 0$ corresponds to nonzero anomalous averages

$$F(\mathbf{K}, \mathbf{k}) = \left\langle \hat{a}_{\frac{\mathbf{K}}{2}-\mathbf{k}\downarrow} \hat{a}_{\frac{\mathbf{K}}{2}+\mathbf{k}\uparrow} \right\rangle, \quad (1)$$

where $\hat{a}_{\mathbf{q}\sigma}$ is the annihilation operator of an electron with momentum $\mathbf{q}$ and spin $\sigma = \uparrow, \downarrow$; the averaging is performed at temperature T. The energy gap parameter

$$\Delta(\mathbf{K}, \mathbf{k}) = -N^{-1}\sum_{\mathbf{k}'} U(\mathbf{k}-\mathbf{k}')F(\mathbf{K}, \mathbf{k}'), \quad (2)$$

where $U(\mathbf{k} - \mathbf{k}')$ is the Fourier transform of the energy of pairing interaction and N is the number of unit cells, is a solution (up to a phase factor) to the self-consistency equation

$$\Delta(\mathbf{K}, \mathbf{k}) = \frac{-1}{2N}\sum_{\mathbf{k}'} U(\mathbf{k}-\mathbf{k}')\Delta(\mathbf{K}, \mathbf{k}')\, f(\mathbf{K}, \mathbf{k}';\, T), \quad (3)$$

where

$$f(\mathbf{K}, \mathbf{k};\, T) = E^{-1}\tanh\frac{E}{2T}, \quad (4)$$

$E = \sqrt{\xi^2 + |\Delta|^2}$ is the energy of the quasiparticle, $\Delta \equiv \Delta(\mathbf{K}, \mathbf{k})$,

$$2\xi \equiv 2\xi(\mathbf{K}, \mathbf{k}) = \varepsilon(\mathbf{K}/2 + \mathbf{k}) + \varepsilon(\mathbf{K}/2 - \mathbf{k}) - 2\mu \quad (5)$$

is the kinetic energy of the relative motion of the pair with total momentum $\mathbf{K}$ measured from the chemical potential 2μ, and $\varepsilon(\mathbf{q})$ is the dispersion law of an electron. The summation is performed over a kinematically allowed domain Ξ_K.

In the BCS model, the attraction between electrons in a narrow layer near the Fermi surface is independent of the momentum $\mathbf{k}$ of the relative motion of the pair; therefore, the order parameter does not depend on $\mathbf{k}$ either. If the pairing interaction is repulsive, then it turns out that a nontrivial solution to the self-consistency equation exists provided that the linear integral operator with the kernel $U(\mathbf{k} - \mathbf{k}')$ has at least one negative eigenvalue [24] in addition to positive eigenvalues.

The perfect mirror nesting leads to a logarithmic singularity with respect to $|\Delta|$ in Eq. (3), which in this case has a solution for arbitrarily small values of the effective coupling constant. Deviations from the perfect mirror nesting under the variation of the momentum of the pair or the shape of the FC smooth out the logarithmic singularity. As a result, pairing becomes possible when the coupling constant exceeds a certain minimal value that increases with the deviation [25]. The energy gap $\Delta(\mathbf{K}, \mathbf{k})$ is an alternating function of the momentum of the relative motion within the domain Ξ_K. The nodal line of this function intersects the PFC [24], thus giving rise to nodes that are exclusively attributed to the repulsive character of the interaction. The maximal value of $|\Delta(\mathbf{K}, \mathbf{k})|$ determines the momentum $\mathbf{K} = \mathbf{K}_j$ of the pair that corresponds to the condensate (the index j numbers crystallographically equivalent momenta).

The eigenfunctions $\varphi_{js}(\mathbf{k})$ of the Hermitian operator with kernel $U(\mathbf{k} - \mathbf{k}')$ form a complete orthonormal system of functions in a domain Ξ_j, and the energy gap can be expanded in this system,

$$\Delta(\mathbf{K}_j, \mathbf{k}) = \sum_s \Delta_s(\mathbf{K}_j)\varphi_{js}(\mathbf{k}); \quad (6)$$

therefore, the dependence of the energy gap on the momentum of relative motion is transferred to the eigenfunctions that are independent of the self-consistency equation. For large $\mathbf{K}_j$, the domain Ξ_j is rather small; therefore, the kernel $U(\mathbf{k} - \mathbf{k}')$ can be expanded in a power series within this domain. If we restrict ourselves to the first two terms of this series [24], we obtain

$$U(\mathbf{k} - \mathbf{k}') \approx U_0 r_0^2[1 - r_0^2|\mathbf{k} - \mathbf{k}'|^2/2], \quad (7)$$

where U_0 and r_0 have the meaning of the effective coupling constant and the screening length, respectively. Kernel (7) is the simplest degenerate kernel with two even and two odd (with respect to the transformation $\mathbf{k} \longleftrightarrow -\mathbf{k}$) eigenfunctions [26].

Owing to the crystalline symmetry, every state of a pair with momentum $\mathbf{K}_j \neq 0$ is degenerate and is formed by a linear combination of pairs with equivalent

momenta. The momentum of relative motion of such a pair belongs to the union of kinematically allowed domains that correspond to different equivalent $\mathbf{K}_j$. The energy gap is rewritten as

$$\Delta(\mathbf{k}) = \sum_j \beta_j \Delta(\mathbf{K}_j, \mathbf{k}) = \sum_{j,s} \beta_j \Delta_s(\mathbf{K}_j)\varphi_{js}(\mathbf{k}), \quad (8)$$

where $\mathbf{k}$ belongs to a certain domain Ξ_j in each term and the coefficients $\Delta_s(\mathbf{K}_j) \equiv \Delta_s$ are equal for all equivalent Ξ_j. The coefficients β_j (normalized by the condition $\sum_j |\beta_j|^2 = 1$) are defined by an irreducible (not necessarily trivial) representation of a symmetry group according to which the functions (8) are transformed and which depends on the interaction that mixes the momenta of the relative motion of pairs with different $\mathbf{K}_j$. If either kinematically allowed domains with different $\mathbf{K}_j$ do not overlap or this overlapping can be neglected, then the squared modulus of the linear combination (8), which can be considered as the order parameter, is represented as

$$|\Delta(\mathbf{k})|^2 = \sum_{j,s,s'} |\beta_j|^2 \Delta_s^* \Delta_{s'} \varphi_{js}^*(\mathbf{k})\varphi_{js'}(\mathbf{k}); \quad (9)$$

this expression allows one to expand the free energy in powers of $|\Delta(\mathbf{K}_j, \mathbf{k})|^2$ for a certain $\mathbf{K}_j$ and then pass to (8).

3. At temperatures close to T_C, $|\Delta(\mathbf{K}_j, \mathbf{k})| \longrightarrow 0$. This allows one to linearize the self-consistency equation for $\tau \equiv (T_C - T)/T_C \ll 1$, representing it as

$$\sum_{\mathbf{k}'} P_{\mathbf{k}\mathbf{k}'}(\mathbf{K}_j; T_C)\Delta(\mathbf{K}_j, \mathbf{k}') = 0, \quad (10)$$

where

$$P_{\mathbf{k}\mathbf{k}'}(\mathbf{K}; T) = NU^{-1}(\mathbf{k} - \mathbf{k}') + \delta_{\mathbf{k}\mathbf{k}'} f_0(\mathbf{K}, \mathbf{k}; T)/2, \quad (11)$$

$f_0(\mathbf{K}, \mathbf{k}; T)$ is the function (4) for $\Delta = 0$ and $U^{-1}(\mathbf{k} - \mathbf{k}')$ is the inverse matrix of $U(\mathbf{k} - \mathbf{k}')$. Using the expansion (6) of the energy gap in the eigenfunctions of the kernel (7), one can reduce the system of equations (10) to a system of two equations and determine the transition temperature from the solvability condition of the latter system. Note that, when the kernel has a unique eigenfunction that corresponds to a positive eigenvalue (an approximation similar to the BCS model for repulsive interaction), the only equation obtained from (10) has only the trivial solution $\Delta = 0$. The spectrum of the kernel (7) contains one negative eigenvalue; therefore, the self-consistency equation certainly has a nontrivial solution [24] for $T < T_C \neq 0$.

For $T < T_C$, the terms with the momenta $\mathbf{K}_j$ of the condensate in the linearized Hamiltonian give rise to a macroscopically large number of pairs. For $T \longrightarrow T_C$, the contribution of these terms can be considered as a small perturbation of the Hamiltonian of an ideal gas of pairs. The free energy can be calculated within thermodynamic perturbation theory [27]; in the second order, it is expressed as

$$F^{(2)} = -\sum_{j,\mathbf{k},\mathbf{k}'} \Delta^\dagger(\mathbf{K}_j, \mathbf{k}) P_{\mathbf{k}\mathbf{k}'}(\mathbf{K}_j; T)\Delta(\mathbf{K}_j, \mathbf{k}'), \quad (12)$$

where the kernel (11) is determined at a temperature T close to T_C. Formula (12) can be rewritten as

$$F^{(2)} = -\sum_{j,\mathbf{k}} \Delta^\dagger(\mathbf{K}_j, \mathbf{k}) P(\mathbf{k}; T)\Delta(\mathbf{K}_j, \mathbf{k}), \quad (13)$$

where the kernel

$$P(\mathbf{k}; T) \approx \frac{\tau}{4T_C} \cosh^{-2}\left[\frac{\xi(\mathbf{K}_j, \mathbf{k})}{2T_C}\right] \quad (14)$$

with $k \in \Xi_j$ is identical for all Ξ_j. Using expansion (6), we obtain

$$F^{(2)} = -\tau \sum_{s,s'} \Delta_s^* F_{ss'} \Delta_{s'}, \quad (15)$$

where the matrix elements

$$F_{ss'} \equiv \frac{1}{4T_C} \sum_{\mathbf{k}} \frac{\varphi_{js}^*(\mathbf{k})\varphi_{js'}(\mathbf{k})}{\cosh^2[\xi(\mathbf{K}_j, \mathbf{k})/2T_C]} \quad (16)$$

are independent of j because of the summation over $\mathbf{k}$. The sum over j is included (together with the coefficients β_j) in the definition of Δ_s according to (8).

The fourth-order contribution $F_c^{(4)}$ can be determined from the expression $F = \overline{H} - TS$, where $\overline{H}$ is the mean value of the linearized Hamiltonian. Omitting the calculations presented in [28], we write the final result:

$$F^{(4)} = \frac{1}{2} \sum_{s,s',t,t'} \Delta_s^* \Delta_{s'}^* F_{ss'tt'} \Delta_t \Delta_{t'}. \quad (17)$$

Here, we set $T_n = \pi T_C(2n + 1))$ in the sum

$$F_{ss'tt'} = \sum_{n=0}^{\infty} \sum_{\mathbf{k}} \frac{2T_C \varphi_{js}^*(\mathbf{k})\varphi_{js'}^*(\mathbf{k})\varphi_{jt}(\mathbf{k})\varphi_{jt'}(\mathbf{k})}{[T_n^2 + \xi^2(\mathbf{K}_j, \mathbf{k})]^2} \quad (18)$$

over the union of all Ξ_j and used the well-known repre-

sentation [29]

$$f_0(\mathbf{K}_j, \mathbf{k}; T) = 4T\sum_{n=0}^{\infty}\frac{1}{[\pi T(2n+1)]^2 + \xi^2(\mathbf{K}_j, \mathbf{k})}. \quad (19)$$

4. To take into account the dependence of the order parameter on the radius vector **R** of the center of mass of a pair, we have to pass to the Fourier transform

$$\Delta(\mathbf{R}, \mathbf{k}) = N^{-1/2}\sum_{\mathbf{K}}\Delta(\mathbf{K}, \mathbf{k})\exp(-i\mathbf{K}\cdot\mathbf{R}), \quad (20)$$

where **k** belongs to the domain Ξ_K corresponding to the momentum **K**. In the case of a weakly inhomogeneous system, when the momenta **K** slightly differ from the momentum $\mathbf{K}_j$ of the condensate of a spatially homogeneous state, $\Delta(\mathbf{K}, \mathbf{k})$ can be expanded in the system of functions $\varphi_{js}(\mathbf{k})$ defined in Ξ_j. Then,

$$\Delta_{js}(\mathbf{R}) = \sum_{\mathbf{k}}\Delta(\mathbf{R}, \mathbf{k})\varphi^*_{js}(\mathbf{k}). \quad (21)$$

For a condensate with momentum $\mathbf{K}_j$ in a homogeneous system, we put $\Delta(\mathbf{K}, \mathbf{k}) = \sqrt{N}\,\Delta(\mathbf{K}_j, \mathbf{k})\delta_{\mathbf{K}\mathbf{K}_j}$, so that

$$\Delta_{js}(\mathbf{R}) = \Delta_s\exp(-i\mathbf{K}_j\cdot\mathbf{R}) \quad (22)$$

is similar to the FFLO state [13, 14], the essential difference being that the energy gap under repulsive interaction has two FFLO components (22) coupled by the self-consistency equation. Functions (22) with different *j* are orthogonal to each other:

$$\sum_{\mathbf{R}}\Delta^*_{js}(\mathbf{R})\Delta_{j's'}(\mathbf{R}) = N\Delta^*_s\Delta_{s'}\delta_{jj'}.$$

In a spatially inhomogeneous state, the coefficients Δ_s are functions of **R** that slowly vary on the atomic scale. In this case, different **R** correspond to generally different condensate momenta; therefore, putting, $\mathbf{K} = \mathbf{K}_j + \mathbf{q}$, one should sum over **q** for every *j*. Since the homogeneous contribution to the free energy has already been obtained, the second-order gradient contribution can be calculated for $T = T_C$ and represented as

$$F_g^{(2)} = -\sum_{j,\mathbf{q},\mathbf{k}}\Delta^\dagger(\mathbf{K}_j + \mathbf{q}, \mathbf{k})P'(\mathbf{q}, \mathbf{k})\Delta(\mathbf{K}_j + \mathbf{q}, \mathbf{k}), \quad (23)$$

where

$$P'(\mathbf{q}, \mathbf{k}) = [f_0(\mathbf{K}_j + \mathbf{q}, \mathbf{k}; T_C) - f_0(\mathbf{K}_j, \mathbf{k}; T_C)]/2N, \quad (24)$$

because, for small **q**, one can apply Eq. (10) to transform the kernel in (12).

The application of a magnetic field changes the phase of the annihilation operator of an electron at a site **n** [30],

$$\hat{c}_{\mathbf{n}\sigma} \longrightarrow \hat{c}_{\mathbf{n}\sigma}\exp\left[\frac{ie}{\hbar c}\mathbf{A}(\mathbf{n})\cdot\mathbf{n}\right]; \quad (25)$$

therefore, the phase of the order parameter as a function of the vector potential **A** is expressed as

$$\frac{e}{\hbar c}[\mathbf{A}(\mathbf{n})\cdot\mathbf{n} + \mathbf{A}(\mathbf{n}')\cdot\mathbf{n}'] \approx \frac{2e}{\hbar c}\mathbf{A}(\mathbf{R})\cdot\mathbf{R}, \quad (26)$$

where $2\mathbf{R} = \mathbf{n} + \mathbf{n}'$, and the approximate equality holds if the derivatives of the vector potential are negligible (in the weak-field approximation). The vector potential **A** in (25) includes both the contribution of a slowly varying field that is retained in the approximate expression (26) and the contribution of the internal field that arises due to doping and current circulations [6]. This contribution manifests itself, for instance, in the boson model [31] of resonating valence bonds [32], as gauge fields that link the charge and spin degrees of freedom.

The energy gap can also be expanded in terms of the eigenfunctions of the kernel of the interaction operator for momenta that differ from the condensate momentum for a spatially homogeneous state. In this case, for small enough **q**, the functions $\varphi_s(\mathbf{k})$ can be assumed to belong to the domain Ξ_j that corresponds to the momentum $\mathbf{K}_j$. Define the functions

$$\Psi_s(\mathbf{R}) = \Delta_s(\mathbf{R})/aT_C, \quad (27)$$

where

$$\Delta_s(\mathbf{R}) = \sum_j\beta_j e^{i\mathbf{K}_j\cdot\mathbf{R}}\Delta_{js}(\mathbf{R}). \quad (28)$$

The normalization condition for the dimensionless functions $\varphi_{js}(\mathbf{k})$ implies that $\varphi_{js}(\mathbf{k}) \sim N_c^{-1/2}$, where $N_c \equiv \Xi_j S/(2\pi)^2$ is the number of states in the domain Ξ_j and *S* is a normalizing area. Then, formulas (15)–(18) show that the order parameter is normalized in the domain of kinematic constraint $|\Psi_s(\mathbf{R})|^2 \sim N_c$, so that $\Psi_s(\mathbf{R})$ can be considered as components of the condensate wavefunctions. The fact that the superfluid density is proportional to N_c for a large total momentum of a pair indicates that the domain of the momentum space in which the scattering of pairs due to the interaction occurs is determined by the domain of kinematic constraint. This contrasts with the BCS model, in which, in spite of the fact that the pairing interaction is different from zero in a thin layer enclosing the Fermi surface, the domain of the momentum transfer under scattering is the entire

Brillouin zone, so that the superfluid density is proportional to the total number of particles. Formula (23) is rewritten as

$$F_{\rm g}^{(2)} = -\frac{1}{N}\sum_{j,\mathbf{q}}\sum_{s,s'}\sum_{\mathbf{R},\boldsymbol{\rho}}\Psi_s^*(\mathbf{R})Q_{ss'}(\mathbf{q})\Psi_{s'}(\mathbf{R}+\boldsymbol{\rho}) \times e^{i\mathbf{q}\cdot\boldsymbol{\rho}}\exp\left\{-i\frac{2e}{\hbar c}\mathbf{A}\cdot\boldsymbol{\rho}\right\}. \quad (29)$$

Here,

$$Q_{ss'}(\mathbf{q}) = a^2T_{\rm C}^2\sum_{\mathbf{k}}\varphi_s^*(\mathbf{k})P'(\mathbf{q},\mathbf{k})\varphi_{s'}(\mathbf{k}), \quad (30)$$

where $a^2 = S/N$.

The "biased" order parameter is expressed as

$$\Psi_{s'}(\mathbf{R}+\boldsymbol{\rho})\exp\left\{-i\frac{2e}{\hbar c}\mathbf{A}\cdot\boldsymbol{\rho}\right\} = e^{i\boldsymbol{\rho}\cdot\hat{\mathbf{D}}}\Psi_{s'}(\mathbf{R}), \quad (31)$$

where the operator of covariant differentiation with respect to the coordinates of the center of mass of a pair is given by

$$\hat{\mathbf{D}} = -i\nabla - \frac{2e}{\hbar c}\mathbf{A}. \quad (32)$$

Up to the second order in $\mathbf{R}$, formula (29) can be represented as (summation over repeated Cartesian indices α and β is assumed)

$$F_{\rm g}^{(2)} = \int d^2R\sum_{s,s'}[\hat{D}_\alpha\Psi_s(\mathbf{R})]^\dagger Q_{ss'}^{\alpha\beta}[\hat{D}_\beta\Psi_{s'}(\mathbf{R})], \quad (33)$$

where, upon passing to integration with respect to $\mathbf{q}$ and $\boldsymbol{\rho}$, we have

$$Q_{ss'}^{\alpha\beta} = a^2T_{\rm C}^3\sum_{n=0}^{\infty}\int\frac{d^2k}{(2\pi)^2}\varphi_s^*(\mathbf{k})\varphi_{s'}(\mathbf{k}) \times\int\frac{d^2\rho}{(2\pi)^2}\rho_\alpha\rho_\beta\int\frac{d^2q}{(2\pi)^2}\frac{e^{i\mathbf{q}\cdot\boldsymbol{\rho}}}{T_n^2+\xi^2(\mathbf{K}_j+\mathbf{q};\mathbf{k})}. \quad (34)$$

The integral with respect to $\mathbf{q}$ is determined by the poles of the integrand in the complex plane (for each component of the vector $\mathbf{q}$), which are obtained from the equations

$$\xi(\mathbf{K}_j+\mathbf{q},\mathbf{k}) = \pm iT_n. \quad (35)$$

For a weakly inhomogeneous state (for small $\mathbf{q}$), the left-hand side of Eq. (35) can be expanded in series in $\mathbf{q}$. Restricting the expansion to the first nonvanishing terms, we obtain

$$\xi(\mathbf{K}_j+\mathbf{q},\mathbf{k}) = \xi(\mathbf{K}_j,\mathbf{k}) + \frac{\hbar^2}{8m^*}a_{\alpha\beta}q_\alpha q_\beta. \quad (36)$$

There are no terms linear in $\mathbf{q}$ in this formula because the electron velocities at the points $\mathbf{k}_+$ and $\mathbf{k}_-$, which are symmetric with respect to $\mathbf{K}_j$, have opposite directions. m^* makes the sense of a certain effective mass of an electron, and the tensor $a_{\alpha\beta}$, defined by the equation

$$a_{\alpha\beta} = \frac{m^*}{\hbar^2}\left[\frac{\partial^2\varepsilon(\mathbf{k}_+)}{\partial k_\alpha\partial k_\beta} + \frac{\partial^2\varepsilon(\mathbf{k}_-)}{\partial k_\alpha\partial k_\beta}\right], \quad (37)$$

can be assumed to be independent of $\mathbf{k}$.

The disposition of the PFC in the extended neighborhood of a saddle point [33] allows one to restrict the analysis to the case of extremely strong anisotropy of the effective masses. Let m^* be a light component of the effective mass, which corresponds to the momentum $\mathbf{K}_j$ of the pair directed along the coordinate axis q_1. Set $a_{\alpha\beta}q_\alpha q_\beta = q_1^2$. Then, integration with respect to q_2 leads to the substitution $\rho_\alpha\rho_\beta \longrightarrow \rho_1^2\delta_{\alpha 1}\delta_{\beta 1}$ in (34). The integral with respect to q_1 can be calculated by the theorem of residues. Subsequent integration with respect to $\boldsymbol{\rho}$ allows us to represent the gradient contribution to the free energy as

$$F_{\rm g}^{(2)} = \frac{\hbar^2}{4m}\int d^2R\sum_{s,s'}[\hat{\mathbf{D}}\Psi_s]^\dagger M_{ss'}[\hat{\mathbf{D}}\Psi_{s'}], \quad (38)$$

where $m = |m^*|$, $\Psi_s \equiv \Psi_s(\mathbf{R})$, and

$$M_{ss'} = 8\sum_{n=0}^{\infty}\int\frac{d^2k}{(2\pi)^2}\frac{a^2T_C^3\varphi_s^*(\mathbf{k})|\xi(\mathbf{K}_j,\mathbf{k})|\varphi_{s'}(\mathbf{k})}{[T_n^2+\xi^2(\mathbf{K}_j,\mathbf{k})]^2}; \quad (39)$$

here we took into account that the momenta $\pm\mathbf{K}_j$ make equal contributions to (34) and that, in the case of tetragonal symmetry, the momenta equivalent to $\pm\mathbf{K}_j$, turned through angles of $\pm\pi/2$, give similar contributions with a simultaneous change of the numbering of the coordinate axes.

Homogeneous contributions of the second (15) and fourth (17) orders are expressed as

$$F^{(2)} = -\tau\int d^2R\sum_{s,s'}\Psi_s^*A_{ss'}\Psi_{s'}, \quad (40)$$

and

$$F^{(4)} = \frac{1}{2}\int d^2R \sum_{s,s',t,t'} \Psi_s^*\Psi_{s'}^* B_{ss'tt'}\Psi_t\Psi_{t'}, \qquad (41)$$

respectively, where

$$A_{ss'} = \frac{a^2 T_C}{4}\int \frac{d^2k}{(2\pi)^2}\frac{\varphi_{js}^*(\mathbf{k})\varphi_{js'}(\mathbf{k})}{\cosh^2[\xi(\mathbf{K}_j,\mathbf{k})/2T_C]} \qquad (42)$$

and

$$B_{ss'tt'} = 2a^4 T_C^5 \times \sum_{n=0}^{\infty}\int \frac{d^2k}{(2\pi)^2}\frac{\varphi_{js}^*(\mathbf{k})\varphi_{js'}^*(\mathbf{k})\varphi_{jt}(\mathbf{k})\varphi_{jt'}(\mathbf{k})}{[T_n^2+\xi^2(\mathbf{K}_j,\mathbf{k})]^2} \qquad (43)$$

in accordance with (16) and (18). Note that, in the simplest case (which is worth considering only for the attractive interaction) when the degenerate kernel of the interaction operator has a unique eigenfunction $\varphi_{js} = N_c^{-1/2}$, the matrices introduced by formulas (39), (42), and (43) degenerate into the numbers

$$M = \frac{gT_C a^2}{N_c}, \quad A = \frac{ga^2T_C^2}{N_c}, \quad B = \frac{7\zeta(3)}{8\pi^2}\frac{ga^4T_C^2}{N_c^2}, \qquad (44)$$

respectively, where g is the density of states and $\zeta(z)$ is the Riemann zeta function.

Let us represent the free energy as the Ginzburg–Landau functional

$$F = F_g^{(2)} + F^{(2)} + F^{(4)} + F^{(m)},$$

where

$$F^{(m)} = \frac{z_0}{8\pi}\int d^2R(\mathrm{curl}\mathbf{A})^2 \qquad (45)$$

is the magnetic field energy, $\mathbf{A} \equiv \mathbf{A}(\mathbf{R})$, and z_0 is the distance between adjacent CuO_2 planes.

A standard variational procedure yields a system of equations for the order parameter,

$$\frac{\hbar^2}{4m}\sum_{s'} M_{ss'}\left[-i\nabla - \frac{2e}{\hbar c}\mathbf{A}\right]\Psi_{s'} - \tau\sum_{s'} A_{ss'}\Psi_{s'} + \sum_{s',t,t'} B_{ss'tt'}\Psi_{s'}^*\Psi_t\Psi_{t'} = 0, \qquad (46)$$

and an equation for the vector potential,

$$\mathbf{j} = \sum_{s,s'} M_{ss'}\left[\frac{\hbar e}{2im}(\Psi_s^*\nabla\Psi_{s'} - \Psi_{s'}\nabla\Psi_s^*) - \frac{2e^2}{mc}\Psi_s^*\Psi_{s'}\mathbf{A}\right], \qquad (47)$$

where

$$\mathbf{j} = z_0\frac{c}{4\pi}\mathrm{curl}\,\mathrm{curl}\mathbf{A} \qquad (48)$$

is the current density in the CuO_2 plane. The system of boundary conditions can be expressed, in particular, as

$$\sum_{s'} M_{ss'}\left[i\nabla + \frac{2e}{\hbar c}\mathbf{A}\right]\Psi_{s'}\mathbf{n} = 0, \qquad (49)$$

where $\mathbf{n}$ is the outward normal to the boundary of a 2D domain in the CuO_2 plane.

The presence of the system of two equations (46) instead of a single Ginzburg–Landau equation corresponding to the BCS model may lead, just as in the case of *s–d* pairing [34], to several nontrivial solutions (that may differ, for example, in the relative phase). These solutions correspond to the minima of the Ginzburg–Landau functional, whose positions and the energy values depend on the relation between the matrix elements $A_{tt'}$ and $B_{stt's'}$. Moreover, system of equations (46) for the interrelated components of the order parameter may lead to topological defects in the phase of the order parameter that are different from vortices and antivortices and arise as solutions to the Ginzburg–Landau equations corresponding to the BCS model, like in the case described by the two-component Gross–Pitaevskii equation for condensates of oppositely charged particles coupled through the electromagnetic field alone [35].

The components of the equilibrium order parameter in a spatially homogeneous system are determined by the system of equations

$$\sum_{s'}\Psi_{s'}\left[\tau A_{ss'} - \sum_{t,t'} B_{stt's'}\Psi_t^*\Psi_{t'}\right] = 0, \qquad (50)$$

which implies that, for $\tau < 0$, the equilibrium state corresponds to the trivial solution $\overline{\Psi}_s = 0$. When $\tau > 0$, the equilibrium order parameter is determined from the condition that the matrix in square brackets in (50) vanish. Denote $\overline{\Psi}_s = \sqrt{\tau}\,c_s$, where the coefficients c_s are independent of temperature and are determined form the system of three independent equations

$$\sum_{s,s'} B_{tss't'}c_s^* c_{s'} = A_{tt'}. \qquad (51)$$

The coefficients c_s enter this system as three nonnegative invariant combinations: as the squared moduli $|c_1|^2$ and $|c_2|^2$ and as the interference term $c_1^* c_2 + c_1 c_2^*$, which determines the relative phase of the coefficients c_1 and c_2. The matrices $A_{tt'}$ and $B_{tss't'}$ have three and five independent components, respectively, which can easily be estimated. Let us choose the eigenfunction φ_{j2}, which corresponds to a positive eigenvalue, in the form that was used for estimates (44). Then, the function φ_{j1}, which corresponds to a negative eigenvalue and vanishes on the PFC, can be chosen as a linear function of ξ. This choice is consistent with the assumption [36] about the "tilt" of the energy gap. For such eigenfunctions, the elements of the matrices $A_{tt'}$, $B_{tss't'}$, and $M_{tt'}$ all of whose indices are 2 coincide with estimates (44) in order of magnitude. Taking into account that the parameter $\gamma = \pi T_C/\varepsilon_0$ is small, we can estimate the above matrix elements with all indices equal to unity as $A_{11} \sim \gamma^2 A$, $B_{1111} \sim \gamma^2 B$, and $M_{11} \sim \gamma M$. $B_{1122} \sim \gamma^2 B$ to a logarithmic degree of accuracy in γ, and the elements with an odd number of identical indices, which are also proportional to γ^2, have an additional small term (which determines their sign) associated with the asymmetry between the filled and vacant parts of the domain Ξ_j.

Note that, when the kernel of the interaction operator has a single positive eigenvalue, system (50) degenerates into a single equation that has no nontrivial solutions. Moreover, system (50) has only the trivial solution when all eigenvalues of the kernel are nonnegative. In the case of two (one positive and one negative) eigenvalues, which is considered here, the elements of the matrices $A_{tt'}$ and $B_{stt's'}$ are such that a nontrivial solution certainly exists for $\tau > 0$.

The essential difference between the system of equations (46) and the system that describes, for example, the *s*–*d* pairing [34], is that the coefficients multiplying the first-order components of the order parameter are proportional to τ. In the case of repulsion, both components Ψ_1 and Ψ_2 correspond to the same transition temperature and cannot exist one without other [24] because the transition to the limit case of a unique positive eigenvalue implies the simultaneous transition $T_C \longrightarrow 0$ [24].

5. It follows from (47) and (48) that the penetration depth λ is defined as

$$\frac{1}{\lambda^2} = \frac{8\pi e^2 \tau}{mc^2 z_0} \sum_{s,s'} c_s^* M_{ss'} c_{s'} \tag{52}$$

and the two-dimensional superfluid density is given by

$$n_s = 2\tau \sum_{s,s'} c_s^* M_{ss'} c_{s'}, \tag{53}$$

where the coefficients c_s are solutions to the system of equations (51).

The size of a pair in the real space can be evaluated as the inverse of the imaginary part of the pole defined by (35) for $n = 0$. Let us denote by ε_0 the characteristic energy scale of the domain Ξ_j and take into account that $\varepsilon_0 \gg T_C$. Then, we obtain the estimate

$$\zeta_0 \approx \frac{\hbar}{\pi T_C} \sqrt{\frac{\varepsilon_0}{|m^*|}}. \tag{54}$$

Setting $\varepsilon_0 \sim 0.1$ eV, we obtain $\zeta_0 \approx 3 \times 10^{-7}$ cm for $|m^*| \approx 10^{-27}$ g and $T_C \approx 80$ K. Note that, for the pairing with zero total momentum (in the case of attraction), when there are no kinematic constraints and ε_0 has the sense of the Fermi energy, we have $\sqrt{\varepsilon_0/|m^*|} \sim v_F$, and (54) is rewritten as $\zeta_0 \sim \hbar v_F/\pi T_C$.

The correlation function for the fluctuations of the components of the order parameter can be represented as

$$\langle \Psi_s^*(\mathbf{R}) \Psi_{s'}(\mathbf{R}') \rangle = \frac{mT_C}{\hbar^2} \sum_t U_{s't} M_t^{-1} U_{ts}^{-1} K_0\left(\frac{r}{r_t^{(c)}}\right), \tag{55}$$

where $\mathbf{r} = \mathbf{R}' - \mathbf{R}$, $U_{ss'}$ is a unitary matrix that simultaneously diagonalizes the matrices $A_{ss'}$ and $M_{ss'}$, A_t and M_t are the eigenvalues of these matrices, and

$$r_t^{(c)} = \sqrt{\frac{\hbar^2}{4m|\tau|} \frac{M_t}{A_t}}. \tag{56}$$

Taking into account that the Macdonald function $K_0(z)$ behaves as

$$K_0(z) \sim (\pi/2z)^{1/2} \exp(-z)$$

for large values of the argument, we can consider the least of quantities (56), $r_c = \min(r_t^{(c)})$, as the characteristic correlation length for the fluctuations of the order parameter. Just like the size of a pair in the real space, this length is determined by the electron effective mass near the FC.

As rough estimates, one can use expressions (44), which are the greatest (in order of magnitude) diagonal elements of appropriate matrices corresponding to the positive eigenvalue of degenerate kernel (7). The mean squared modulus of the equilibrium order parameter is proportional to the area of the domain Ξ_j,

$$\overline{|\Psi_s|^2} \sim [2\tau/7\zeta(3)] N \Xi_j, \tag{57}$$

and the dimensionless ratio $\kappa = \lambda/r_c \gg 1$.

The gradient contribution to the free energy (40) describes long-wavelength fluctuations of the order parameter with respect to the thermodynamic equilibrium value determined by system of equations (50). The structure of order parameter (28) is such that a special current state may arise in the domain of well-developed fluctuations. The character of this state can be determined by Ginzburg–Landau equations (46) and (47). If we single out the phase

$$\Psi_s(\mathbf{R}) = |\Psi_s(\mathbf{R})|e^{i\Phi(\mathbf{R})} \quad (58)$$

of the order parameter, we can rewrite expression (47) for a superconducting current in the absence of magnetic field as

$$\mathbf{j} = \frac{\hbar e}{m}\overline{M}\nabla\Phi. \quad (59)$$

Here,

$$\overline{M} = \sum_{s,s'}\Psi_s^* M_{ss'}\Psi_{s'}, \quad (60)$$

and Ψ_s is given by expression (27). Due to the exponential factors in the definition (28), Ψ_s is a rapidly varying (on the space scale proportional to K_j^{-1}) function of the coordinates X and Y of the radius vector $\mathbf{R}$:

$$\Psi_s(\mathbf{R}) \sim \cos K_j X \pm \cos K_j Y. \quad (61)$$

The upper (lower) sign corresponds to the irreducible representation of A_{1g} (B_{1g}). The current circulation along any closed contour L_0 in the cuprate plane,

$$V = \oint_{L_0}\mathbf{j}\,d\mathbf{l} = \frac{\hbar e}{m}\int_{S_0} d^2R\ \mathbf{n}_3[\nabla\overline{M}\times\nabla\Phi], \quad (62)$$

is generally different from zero; here, $\mathbf{n}_3$ is a unit normal to the CuO_2 plane, and S_0 is the area of the surface bounded by the contour L_0. If the characteristic size of contour L_0 is much greater than the scale proportional to K_j^{-1}, on which a variation of the order parameter (61) occurs, then $V \approx 0$. The structure of order parameter (61) corresponds to the partition of a real 2D space into cells of area on the order of K_j^{-2}, so that the projections of the vector product in (62) onto direction $\mathbf{n}_3$ in adjacent cells have different signs. Thus, long-wavelength fluctuations of the phase of the order parameter (which are especially significant in underdoped cuprates due to the small phase stiffness) result in an incommensurable antiferromagnetically ordered (on the scale of phase fluctuations) structure in the form of orbital current circulations.

An estimate for the phase stiffness near the transition temperature can be obtained from Eq. (38). In the absence of magnetic field, expression (38) can be rewritten as

$$F_g^{(2)} \approx \frac{1}{2}\int d^2R\rho_s(\nabla\Phi)^2, \quad (63)$$

whence it follows that

$$\rho_s = \frac{\hbar^2}{2m}\overline{M}. \quad (64)$$

Here, $\overline{M}$ is defined by (60) in which the components of the order parameter should be set equal to their equilibrium values, which are solutions to system of equations (50). Thus, near the transition temperature, we have $\rho_s \sim \tau$. If we assume that ρ_s is a linear function of temperature [37], then, for $\tau = 1$, expression (64) determines the value of the phase stiffness at zero temperature, $\rho_s(0)$, which is proportional to the area of the domain Ξ_j of kinematic constraint.

6. The mean-field approximation used when deriving Ginzburg–Landau equations (46) and (47) is assumed to be sufficient [1] for describing the SC state of optimally doped and overdoped cuprates. In the underdoped regime, the presence of a pseudogap for $T_C < T < T^*$ suggests [32] that the ground state of the high-temperature ($T > T_C$) phase differs from the state of a normal Fermi liquid. Therefore, a theory based on the same principles as the BCS theory may prove to be inadequate for the physics of cuprates. Moreover, fluctuations play an essential role in underdoped cuprates [4], so that, if the mean value of the order parameter vanishes at a certain transition temperature T_C, its mean square may be different from zero in a rather wide range of temperatures above T_C [3].

Analysis of the t–J model under the constraint of no double occupation of lattice sites shows [18, 38, 39] that the mean-field theory with regard to the relevant renormalization of the kinetic energy and the superexchange interaction in the t–J Hamiltonian gives a reasonable doping-level dependence for the SC-transition temperature T_C and the temperature T^* at which a pseudogap arises. In addition, it shows that the phase stiffness in the underdoped state is small ($\rho_s \sim x$). The latter fact allows one to interpret T^* as the formation temperature of pairs that exist for $T_C < T < T^*$ in noncoherent states, and T_C, as the temperature at which phase coherence arises [18].

The constraint of no double occupation corresponds to the limit case when the energy of on-site correlation is much greater than the electron energy band gap and thus promotes dielectrization of the system and the emergence of AF ordering. A more realistic constraint on the double occupation under which the total (Gutzwiller) projection used in [18, 38, 39] is replaced

by a partial projection [40] also leads to the phase stiffness $\rho_s \sim x$. Moreover, this constraint allows one to apply the general BCS approach to describe SC in underdoped cuprates without resorting (like in [18]) to various, rather factitious schemes for separating charge and spin degrees of freedom [31, 41–43]. The absence of a true separation of charge and spin (a specific confinement of holons and spinons) is an argument in favor of the general BCS conception as applied to cuprates.

Unlike the BCS theory, in which the SC state arises as a result of instability of a normal Fermi liquid with respect to the Cooper pairing with zero total momentum, the theory of superconductivity in cuprates must take into account competition between the SC and AF states [7]. Antiferromagnetic ordering is characterized by a momentum on the order of the doubled Fermi momentum. This is one of the reasons why the SC pairing with large momentum [9–11] (which is generally incommensurate [7] in doped cuprates) may turn out to be the main pairing channel [19].

The description of doped cuprates within the *t–J* model on the basis of SU(2) symmetry in fermionic representation [44] suggests that one should introduce a doublet of bosons one of which may condensate into a state with zero momentum and the other corresponds to a condensate with large momentum (π, π). The SU(2) formalism leads to the correlation of orbital currents, which manifests the AF ordering of current circulations [6] corresponding to a flux phase [45]. Equation (62) implies that such a character of current circulations can be associated directly with the large momentum of the pairs that form the SC condensate.

Owing to the positive term $U_0 r_0^2$, the repulsive pairing interaction (7) takes into account a partial constraint of no double occupation of sites and results in a substantial dependence of the SC order parameter on the momentum of the relative motion of the pair. For large total momentum of the pair, this dependence results, first, from the crystalline symmetry (pairing in equivalent domains Ξ_j of the momentum space) [23, 26], and, second, from the specific character of repulsion (the presence of a negative eigenvalue of the interaction kernel, which is in agreement with the assumption [46] that there is a pairing interaction in addition to the Hubbard repulsion). Thus, the four nodes for the *d*-wave symmetry of the SC gap (the irreducible representation B_{1g} relating equivalent domains Ξ_j) are supplemented with nodes associated with intersection of the FC with the nodal line of the order parameter [24]. Under the *s*-wave symmetry (the trivial representation A_{1g}), the repulsive pairing interaction also gives rise to nodes of the order parameter (the extended *s*-wave symmetry [2]). The analysis [47, 48] of the experimental data obtained by different methods shows [47] that the order parameter in the bulk of a cuprate semiconductor corresponds to the *s*-wave symmetry (which is extended in compounds with hole doping and anisotropic in compounds with electron doping) [47]. This result does not contradict the fact that phase-sensitive experiments that allow one to probe thin (on the order of the coherence length) surface layers indicate that the order parameter has the *d*-wave symmetry [47]. With the nodes of the order parameter and quasiparticles that are thermally generated in the neighborhoods of these nodes, one may associate a linear decrease in the phase stiffness as temperature increases [37]. This makes it possible to determine (within the mean-field theory) the transition temperature from the condition $\rho_s(T_C) = 0$ [37]. Note that the additional nodes, just as the regions of the nodal line of the order parameter that lie near the FC, lead to an increase in the number of generated quasiparticles (compared with, say, the case of "pure" *d*-wave symmetry). Therefore, systems with strong anisotropy of the FC, where the nodal line is rather far from the FC on the average, should be characterized by higher values of T_C.

The smallness of the phase stiffness is responsible for the significant long-wavelength phase fluctuations of the order parameter, which, for $T \approx T_C$, develop into singular fluctuations in the form of uncoupled pairs of vortices and antivortices [4] that exist in the pseudogap state of underdoped cuprates. The Dirac character of the spectrum of nodal quasiparticles [1] makes it possible to describe, by analogy with quantum electrodynamics in two dimensions (QED_{2+1}), the pseudogap state within the phenomenological scheme of [4]. In this scheme, the nonsuperconducting phase with a pseudogap manifests itself as a certain algebraic Fermi liquid [4] that plays the same role in the phase transition to the SC state as the normal Fermi liquid plays in the BCS theory. Vortices and antivortices, which represent elementary excitations in the QED_{2+1} scheme, manifest themselves as topological defects of the phase of the order parameter and result from quantum or thermal fluctuations. Similar defects may arise due to the change of the signs of individual current circulations [49] in a state corresponding to the flux phase. In the energy-band scheme, these defects can be associated with long-lived quasi-stationary states of particle pairs with large total momentum [26].

The phase transition to the SC state in 2D systems in which the order parameter has nodes inevitably acquires the characteristic features of the Berezinskii–Kosterlitz–Thouless transition because quasiparticles reduce the phase stiffness and thus stimulate the thermal excitation of vortices and antivortices. The Josephson junction between CuO_2 layers [18], just as between equivalent domains Ξ_j under pairing with large momentum, may suppress the phase fluctuations and bring the transition temperature closer to the value defined by the mean-field theory. A similar effect results from the competition between the SC and DDW states with regard to the inhomogeneous distribution of charge carriers that are introduced during doping along different copper–oxygen planes in a unit cell [5].

All successful phenomenological approaches to the problem of superconductivity in cuprates are associated with inhomogeneities either in the real space (for example, stripes) or in the momentum space [1]. Pairing with large momentum inevitably results in a spatially inhomogeneous SC state, similar to the FFLO state, in which the distribution of the amplitude of the order parameter is analogous to that in the DDW. The pairing repulsion gives rise to nodes in the order parameter and thus reduces the phase stiffness of the SC condensate. These two principles substantially unify various theoretical approaches to the description of superconductivity in cuprates [50].

ACKNOWLEDGEMENTS

This work was supported in part by the Ministry of Education of the Russian Federation (project no. E02-3.4-147), by the Special Federal Program "Integration" (project no B0049), by the Russian Foundation for Basic Research (project no. 02-02-17133), and by the Special Federal Research Program "Investigations and Development in Priority Fields of Science and Technology" (project nos. 40.072.1.1.1173 and 40.012.1.1.1357).

REFERENCES

1. J. Orenstein and A. J. Millis, Science **288**, 468 (2000).
2. E. W. Carlson, V. J. Emery, S. A. Kivelson, and D. Orgad, in *Physics of Conventional and Unconventional Superconductors*, Ed. by K. H. Bennemann and J. B. Ketterson (Springer, Berlin, 2002).
3. V. J. Emery and S. A. Kivelson, Nature **374**, 434 (2002).
4. M. Franz, Z. Tešanović, and O. Vafek, Phys. Rev. B **66**, 054535 (2002).
5. S. Chakravarty, H.-Y. Kee, and K. Volker, Nature **428**, 53 (2004).
6. D. A. Ivanov, P. A. Lee, and X.-G. Wen, Phys. Rev. Lett. **84**, 3958 (2000).
7. S. Chakravarty, R. B. Laughlin, D. K. Morr, and C. Nayak, Phys. Rev. B **63**, 094503 (2001).
8. P. A. Lee, N. Nagaosa, T. K. Ng, and X.-G. Wen, Phys. Rev. B **57**, 6003 (1998).
9. S.-C. Zhang, Science **275**, 1089 (1997).
10. M. Guidry, L.-A. Wu, Y. Sun, and C.-L. Wu, Phys. Rev. B **63**, 134516 (2001).
11. C. N. Yang, Phys. Rev. Lett. **63**, 2144 (1989).
12. G. I. Japaridze, A. P. Kampf, M. Sekania, *et al.*, Phys. Rev. B **65**, 014518 (2002).
13. P. Fulde and R. A. Ferrel, Phys. Rev. A **135**, 550 (1964).
14. A. I. Larkin and Yu. N. Ovchinnikov, Zh. Éksp. Teor. Fiz. **47**, 1136 (1964) [Sov. Phys. JETP **20**, 762 (1964)].
15. H. Shimahara, cond-mat/0406384.
16. A. A. Abrikosov, Physica C (Amsterdam) **341**–**348**, 97 (2000).
17. E. G. Maksimov, Usp. Fiz. Nauk **170**, 1033 (2000) [Phys. Usp. **43**, 965 (2000)].
18. P. W. Anderson, P. A. Lee, M. Randeria, *et al.*, J. Phys.: Condens. Matter **24**, R755 (2004).
19. V. I. Belyavsky and Yu. V. Kopaev, Phys. Rev. B **67**, 024513 (2003); Phys. Lett. A **322**, 244 (2004).
20. V. L. Ginzburg and L. D. Landau, Zh. Éksp. Teor. Fiz. **20**, 1064 (1950).
21. L. P. Gor'kov, Zh. Éksp. Teor. Fiz. **36**, 1918 (1959) [Sov. Phys. JETP **9**, 1364 (1959)].
22. J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **108**, 1175 (1957).
23. V. I. Belyavskiĭ, V. V. Kapaev, and Yu. V. Kopaev, Zh. Éksp. Teor. Fiz. **118**, 941 (2000) [JETP **91**, 817 (2000)].
24. V. I. Belyavskiĭ, Yu. V. Kopaev, V. M. Sofronov, and S. V. Shevtsov, Zh. Éksp. Teor. Fiz. **124**, 1149 (2003) [JETP **97**, 1032 (2003)]; V. I. Belyavsky, Yu. V. Kopaev, and S. V. Shevtsov, J. Supercond. Novel Magn. **17**, 297 (2004).
25. V. I. Belyavskiĭ, V. V. Kapaev, and Yu. V. Kopaev, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 51 (2002) [JETP Lett. **76**, 44 (2002)].
26. V. I. Belyavskiĭ, Yu. V. Kopaev, Yu. N. Togushova, and S. V. Shevtsov, Zh. Éksp. Teor. Fiz. **126**, 672 (2004) [JETP **99**, 585 (2004)].
27. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics*, 4th ed. (Nauka, Moscow, 1995; Butterworth, London, 1999).
28. V. P. Mineev and K. V. Samokhin, *Introduction to Unconventional Superconductivity* (Mosk. Fiz.–Tekh. Inst., Moscow, 1998; Taylor and Francis, London, 1999).
29. 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).
30. E. M. Lifshitz and L. P. Pitaevskiĭ, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics* (Fizmatgiz, Moscow, 2001; Pergamon, New York, 1980), Part 2.
31. Z. Y. Weng, D. N. Sheng, and C. S. Ting, Phys. Rev. B **59**, 8943 (1999); V. N. Muthukumar and Z. Y. Weng, Phys. Rev. B **65**, 174511 (2002).
32. P. W. Anderson, Science **235**, 1196 (1987).
33. Z.-X. Shen, W. E. Spicer, D. M. King, *et al.*, Science **267**, 343 (1995).
34. Y. Ren, J. H. Xu, and C. S. Ting, Phys. Rev. Lett. **74**, 3680 (1995).
35. E. Babaev, L. D. Faddeev, and A. J. Niemi, Phys. Rev. B **65**, 100512 (2002).
36. J. E. Hirsch, Phys. Rev. B **59**, 11962 (1999); Physica C (Amsterdam) **341**–**348**, 213 (2000).
37. P. A. Lee and X. G. Wen, Phys. Rev. Lett. **78**, 4111 (1997).
38. F. C. Zhang, C. Gros, T. M. Rice, and H. Shiba, Supercond. Sci. Technol. **1**, 36 (1988).

39. A. Paramekanti, M. Randeria, and N. Trivedi, Phys. Rev. Lett. **87**, 217002 (2001); cond-mat/0305611.

40. R. B. Laughlin, cond-mat/0209269.

41. G. Baskaran, Z. Zou, and P. W. Anderson, Solid State Commun. **63**, 973 (1987); P. W. Anderson, G. Baskaran, Z. Zou, and T. Hsu, Phys. Rev. Lett. **58**, 2790 (1987).

42. N. Nagaosa and P. A. Lee, Phys. Rev. Lett. **64**, 2450 (1990).

43. P. B. Wiegmann, Phys. Rev. Lett. **60**, 821 (1988); P. A. Lee, Phys. Rev. Lett. **63**, 680 (1989).

44. X. G. Wen and P. A. Lee, Phys. Rev. Lett. **76**, 503 (1996).

45. J. B. Marston and I. Affleck, Phys. Rev. B **39**, 11538 (1989); T. Hsu, J. B. Marston, and I. Affleck, Phys. Rev. B **43**, 2866 (1991).

46. B. A. Bernevig, G. Chapline, R. B. Laughlin, *et al.*, cond-mat/0312573.

47. G. Zhao, Phys. Rev. B **64**, 024503 (2001).

48. B. H. Brandow, Phys. Rev. B **65**, 054503 (2002).

49. S. Chakravarty, Phys. Rev. B **66**, 224505 (2002).

50. V. I. Belyavskiĭ and Yu. V. Kopaev, Usp. Fiz. Nauk **174**, 457 (2004) [Phys. Usp. **47**, 409 (2004)].

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 100, No. 1, 2005, pp. 50–65.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 1, 2005, pp. 56–73.

ORDER, DISORDER, AND PHASE TRANSITIONS IN CONDENSED SYSTEMS

Resistance–Current Curves of High Pinning Superconductors

E. Yu. Klimenko, A. B. Imenitov, S. V. Shavkin, and P. V. Volkov

Russian Research Center Kurchatov Institute, pl. Kurchatova 1, Moscow, 123182 Russia

e-mail: klimenko_e@mail.ru

Received October 24, 2003; in final form, July 22, 2004

Abstract—The lack of universal phenomenological approaches to calculation of the spatiotemporal distributions of vector fields in superconducting materials with high anisotropic pinning hinders the analysis of accumulated empirical data. The development of such phenomenology is impeded by the tradition of inadequate description of the transition characteristics of superconductors. Based on a detailed investigation of the current-dependent transitions in commercial niobium–titanium wire, it has been found that, in contrast to the commonly accepted approach, the transition characteristics are conveniently described as the dependences of the resistance on the current, by analogy with the dependences on the temperature and magnetic field. Investigations in the range of magnetic fields from 0 to B_{c2} showed that it is the resistance that exhibits exponential variation with the field. The values of resistances determined by extrapolation of the transition curves to a zero current obey the exponential dependence on the magnetic field. The experimental results are consistent with predictions of the statistical model explaining the shape of the transition characteristic as determined by the bulk inhomogeneity of the superconductor. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

After more than four decades of research, the resistive state of superconductors with high pinning is still among the central topics in investigations of the phenomenon of superconductivity. The interest in this issue is related to important technical applications of superconductors capable of carrying nondissipative currents of high densities. It is surprising enough that, despite such a long period of extensive investigation, discoveries are still being made and new physical ideas formulated in this field. The frequency of such events especially increased after the discovery of high-temperature superconductivity and was related to the growing variety of experiments devoted, in particular, to the dependence of electric fields and critical currents on the orientation of an external magnetic field. Each new experimental result was explained by peculiarities of the interaction between magnetic flux quanta and pinning centers of a certain type, which led to a very complicated pattern of interactions of the magnetic flux with the superconductor structure. Now it is difficult to judge unambiguously whether the existing notions reflect the real complexity of the phenomenon under investigation, or if the abundant harvest of discoveries reflects the absence of adequate phenomenology with predictive potential and capable of describing the time variation and spatial distribution of vector fields and currents in a superconductor for arbitrary orientations of external fields and currents relative to the principal axes of the anisotropic structure.

Adhering to the second point of view, we believe that attempts to construct an adequate phenomenological electrodynamics are not futile. Previously, we have formulated such a phenomenology in the critical state approximation [1], which allows one to describe, for arbitrary orientations of the external fields and currents, the static characteristics of a given superconductor (such as the critical current density and the angle between the current density and electric field vectors) in terms of the experimentally determined parameters of the global anisotropic potential well of pinning. The proposed model provides a good description for the results of a large series of experiments performed on a niobium–titanium foil with highly anisotropic pinning. Some predictions made within the framework of this model were confirmed by subsequent experiments [2, 3]. We find it attractive and possible to expand the model and construct a consistent vector electrodynamics of superconductors with high anisotropic pinning, which would describe the process of magnetic flux diffusion in such materials.

In order to ensure that the generalized model would be capable of describing processes developing in time, it is necessary to generalize the approach to description of the transition characteristics of superconductors, which is presently performed only in a scalar form. We believe that good prospects have nothing to do with the well-known attempts [4] to construct a vector electrodynamics of superconductors with high anisotropic pinning based on the concept of a thermally activated vortex motion [5, 6] nor on the generalization of the critical state model $\mathbf{j} = j_c\mathbf{E}/E$ introduced by Carr [7]. The activation model initially employs a scalar expression for the electric field, whereby the absolute value of

a vector product of the current density and the magnetic field enters into an exponent or the hyperbolic sine argument, by no means influencing the direction of the electric field. The Carr approach to generalization is merely incorrect: in fact, the directions of electric field and current even in the isotropic case coincide provided only that the magnetic field is perpendicular to the current direction [1, 2].

An alternative, statistical explanation of the shape of transition characteristics [8] taking into account inhomogeneity of the critical parameters of superconductors can be quite readily generalized. This approach was previously considered as being opposite to that used in the thermal activation model [8–11], but later the regions of applicability of the two models were separated. It was accepted that certain "internal" transition characteristics determined by physical processes (such as thermal activation) are inherent in superconducting materials, whereas the statistical model was invoked in explaining the observed broadening of the transition as caused by the longitudinal inhomogeneity of a superconducting wire [12, 13]. Before the discovery of high-temperature superconductors, it was assumed that there is a certain internal voltage–current curve, which is exponential, while the statistical characteristic obeys a power law. Indeed, model calculations of the transition characteristics of a one-dimensional chain of elements with critical parameters distributed according to the normal law [8, 14, 15] give results obeying a power law in a certain region of the total characteristic. Coincidentally, this region coincided with that experimentally measured for multifilament wires [16]. The voltage–current (V–I) curves of high-temperature superconductors also proved to obey the power law, and the activation model was adapted to this shape of the internal curve by introducing a logarithmic potential well of pinning [17].

Our suggestion made quite long ago [14] concerning application of the statistical model to the internal transition characteristics with due regard of the bulk inhomogeneity of superconductors was not accepted, even despite a convincing, in our opinion, numerical demonstration [15, 18] of the fact that a bulk-inhomogeneous model with a normal distribution of the properties of elements gives an exponential transition characteristic. This result is provided by the possibility of bypassing an element, occurring in the normal state, via superconducting paths. It was shown [18] that the difference between a limited real distribution and the unlimited normal distribution does not significantly influence the shape of the calculated superconducting transition characteristic. Subsequently, numerous experimental data were obtained which confirmed the bulk inhomogeneity of superconducting materials possessing high current-carrying capacity [2, 4, 19–23]. The notion of such inhomogeneity in superconductors with high pinning is now commonly accepted. On the one hand, vortices exhibit pinning on microinhomogeneities; on the other hand, samples frequently comprise more or less ordered mixtures of various superconducting phases. For example, it was reported that several layers of intermetallic phases with various compositions, grain sizes, degrees of straining—and hence differing critical temperatures—were formed on the surface of a niobium filament in a bronze coating [23]. A certain hierarchy of the types of structural inhomogeneities was considered in [19, 23], where it was suggested to place inhomogeneities with dimensions on the order of an alloy grain size between extreme types. The grains in a homogeneous alloy may exhibit a certain scatter in the composition and, hence, critical temperature, electron mean free path and, hence, critical field, concentration of pinning centers and, hence, critical current density. As will be demonstrated below, a statistical distribution of any one critical parameter gives rise to scatter (with the same variance) in the other two such parameters.

Transition characteristics of such inhomogeneous materials are conventionally described using the expression [14, 24–26]

$$\rho_{\text{eff}} = 0.5\rho_n\left\{1 + \tanh\left[-\frac{K(T, B, j)}{2\delta}\right]\right\}, \qquad (1)$$

This formula corresponds to parallel coupling of the normal resistance ρ_n and a superconducting component ρ_{sc} exhibiting exponential growth with the temperature T, magnetic field B, and current density j:

$$\rho_{sc} = \rho_n \exp\left[-\frac{K(T, B, j)}{\delta}\right],$$

where

$$K(T, B, j) = \left(1 - \frac{T}{T_c(0)}\right)\left(1 - \frac{B}{B_{c2}(T)}\right) \times \left(1 - \frac{j}{j^*_{c/2}(T, B)}\right),$$

and δ has the meaning of a relative variance of the distribution of critical parameters in the given inhomogeneous material. The preexponential term in Eq. (1) is selected so that the effective resistance would amount to half of the normal resistance when any parameter reaches its critical value (T_c, B_{c2}, or $j^*_{c/2}$ rendering $K = 0$). This choice corresponds to the adopted definition of critical parameters.

Equation (1) reflects the close relationship between the temperature-, field-, and current-dependent transitions in a superconductor, whereas the thermal activation model considers the current-dependent transition as distinct. The shape of the current-dependent transition characteristic predicted by the activation model

differs from the shape of the field and temperature dependences $R_{\text{eff}}(B, T)$ described by logistic curves. This difference was explained as being related to the fact that the field and temperature are the parameters of state, whereas the current is a kinetic quantity. Interrelation between the temperature-, field-, and current-dependent transitions was usually not discussed. Recently, Hense *et al.* [26] again established a scaling relationship between the temperature-, field-, and current-dependent transitions in a niobium–tin wire. Apparently, it was the inertia of thinking in terms of the thermal activation model that made the authors of [26] conclude that "from the physical point of view, scaling … is not understood at the moment."

Below we present the results of investigations of the transition characteristics of superconducting wires with high pinning. The aim of our study was to provide experimental evidence for the planned continuation of the work on vector electrodynamics of superconducting materials with high pinning. We hope that these results will stimulate revision of some common notions about the nature of the resistive state of such superconductors.

2. EXPERIMENTAL METHODS AND RESULTS

We naturally selected a niobium–titanium wire as the sample material, since it is analogous to niobium–zirconium wires widely used in basic experiments favoring establishment of the thermal activation model [27–30]. However, this is not the only basis for this choice: we believe niobium–titanium wire and foil to be the most suitable materials for investigating the transition characteristics of superconductors with high pinning and studying the problems of their electrodynamics. Commercial technology of niobium–titanium alloys is on a quite high level and provides relatively homogeneous materials exhibiting uniform properties along the wire length, with sufficiently high anisotropic pinning. No other commercially available material offers these advantages, and it is even much more difficult to obtain samples with high longitudinal uniformity under laboratory conditions. These advantages make it possible to study the general laws of electrodynamics in technical superconductors [1], which are almost not masked by specific features of particular samples. Experiments with niobium–titanium wires are not complicated by brittleness and high sensitivity to straining, which we typical of intermetallic compounds and HTSC. On the other hand, niobium–titanium wires are by no means a simple material. These wires were displaced from the focus of research, not even having been exhaustively studied. In recent years, it was found that niobium–titanium alloys are two-component (as manifested by a difference in the critical fields of the grain body and boundaries [2]) and are characterized by anisotropic pinning in the cross section of a round wire [3].

Three U-shaped samples for *V–I* measurements using the standard four-point-probe technique were prepared from a piece of cold-drawn copper coated Nb–50 wt% Ti wire with a diameter of 0.15 mm. The superconducting core diameter was 0.12 mm. The sample ends were soldered to current-carrying leads. Potential tips spaced by 10 mm were connected to a 20-mm-long central part, which was oriented perpendicular to the magnetic field. The magnetic field generated in the working region by a 200-mm-long coil with an inner diameter of 42 mm was homogeneous to within 10^{-4}. This was sufficient to exclude the effect of field inhomogeneity on the *V–I* curve with a rather high slope. In sample 1, the copper coating in the region between potential contacts was removed by etching in an iron chloride solution. The resistance of sample 1 in the normal state was 0.85 Ω. Sample 2, on which the copper coating was retained, had a much smaller resistance (2.1 mΩ). Samples 1 and 2 were fixed with low-temperature glue on a fiberglass substrate. Sample 3, soldered to a stainless steel substrate, had a resistance of 0.56 mΩ. All these normal resistances were determined in the linear regions of the voltage–current curves in high magnetic fields, where the critical currents were very small.

Three differently prepared samples were necessary, on the one hand, for providing satisfactory sensitivity and stability of the *V–I* measurements in liquid helium for current densities ranging within four orders of magnitude and, on the other hand, for excluding burn-out of the samples during the operating time of the sample protecting scheme. This is illustrated in Fig. 1, which shows the nonisothermal *V–I* curves [31] calculated for samples 1–3 in various magnetic fields. Based on some preliminary data concerning properties of the wire studied, we arbitrarily selected the heat transfer coefficients as 10^2 W/(m K) for samples 1 and 2 glued to the substrate and 10^4 W/(m K) for sample 3 with the surface washed by liquid helium. Exact conditions of heat exchange were unknown, the *V–I* measurements were limited from below (on a level of 5 nV) by the intrinsic noise of a photovoltaic nanovoltmeter, and from above, by the development of thermal instability or by the process of current distribution between the superconducting core and the normal-metal shunt. Since a voltage jump corresponding to a vertical tangent to the *V–I* curve takes place upon an insignificant heating [32], the preceding part of the curve remains virtually undistorted. Overheating involved in the distribution of current between the superconductor and the normal metal is also insignificant, which allows the isothermal *V–I* curve of the superconductor to be reconstructed by determining the corresponding current as the difference between the total current and the component passing via the normal metal, which is readily determined for

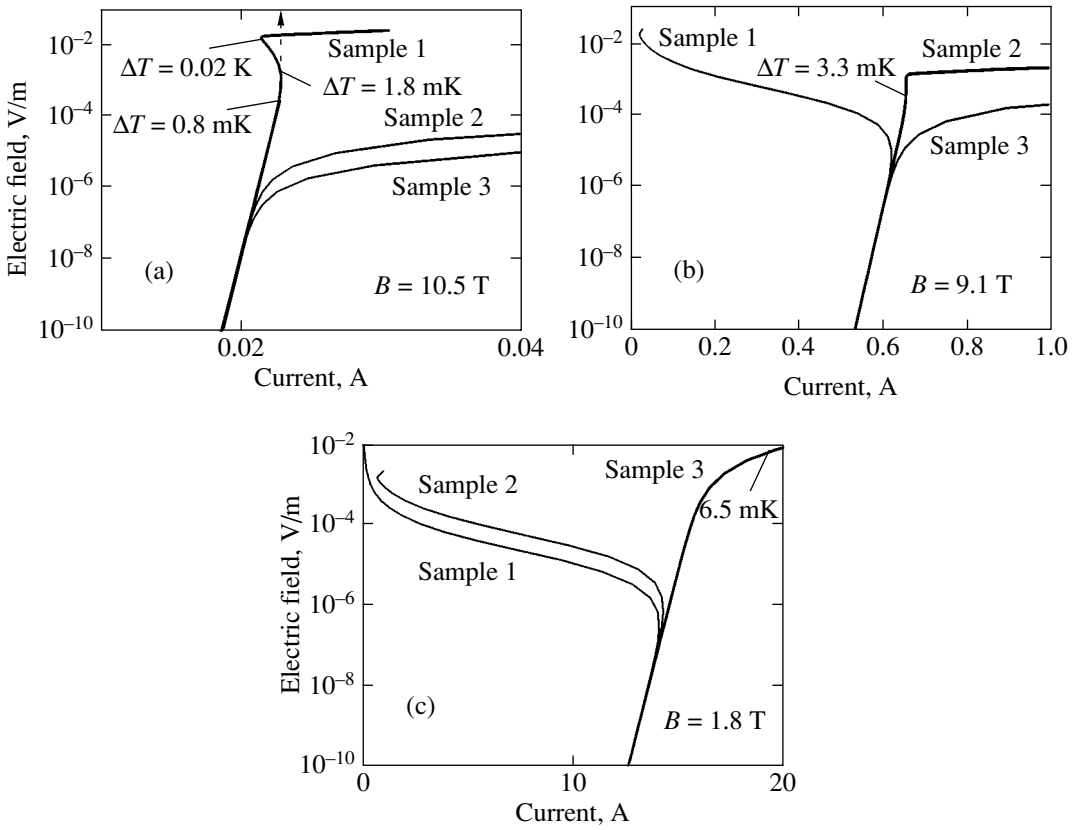


Fig. 1. Voltage–current curves calculated for the niobium-titanium wire samples studied in experiment: (a) the region of fields close to B_{c2} (sample 1 with removed copper coating is stable up to I = 20 mA and ensures obtaining V–I curves with satisfactory precision; the accuracy of measurements on samples 2 and 3 with a stabilizing metal is unsatisfactory); (b) the region of high magnetic fields (sample 1 is unstable; V–I curves can be obtained for sample 2, where the current passing into the copper coating is small, so that the accuracy of its calculation does not significantly affect the determination of current flowing in the superconductor); (c) the region of low magnetic fields (samples 1 and 2 are unstable; V–I curves can be obtained for sample 3, where the current passing into the superconductor is greater than that in the stabilizing substrate and can be calculated with acceptable accuracy).

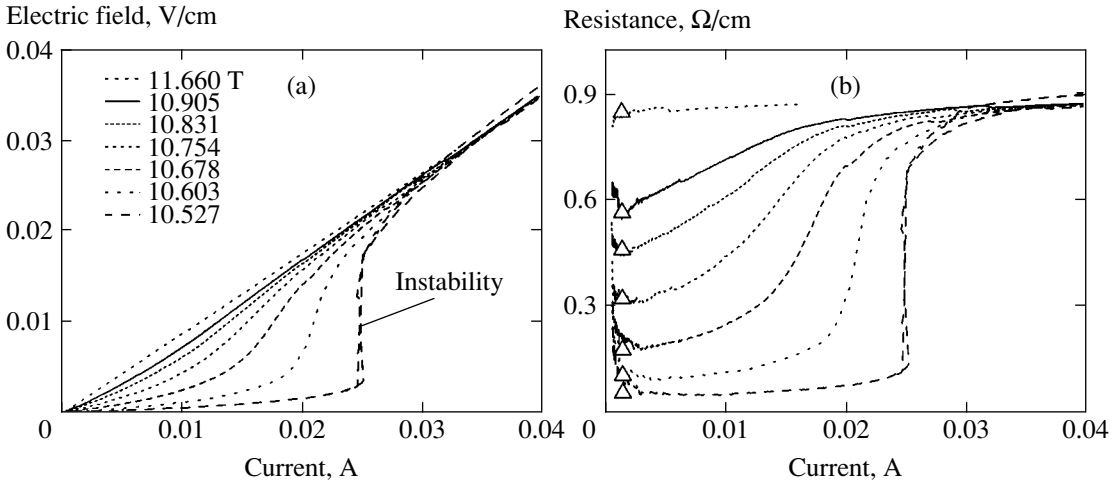


Fig. 2. Experimental (a) voltage–current and (b) resistance–current curves of sample 1 measured in various magnetic fields.

the known resistance. In these calculations, we ignored the field dependence of the substrate resistance for sample 3. The copper coating resistance for sample 2 was determined using the Kohler rule.

As can be seen from Fig. 2a, sample 1 measured at $B > 10.2$ T allowed stable V–I curves of the superconductor to be measured in rather high electric fields. However, it was difficult to use this sample in low

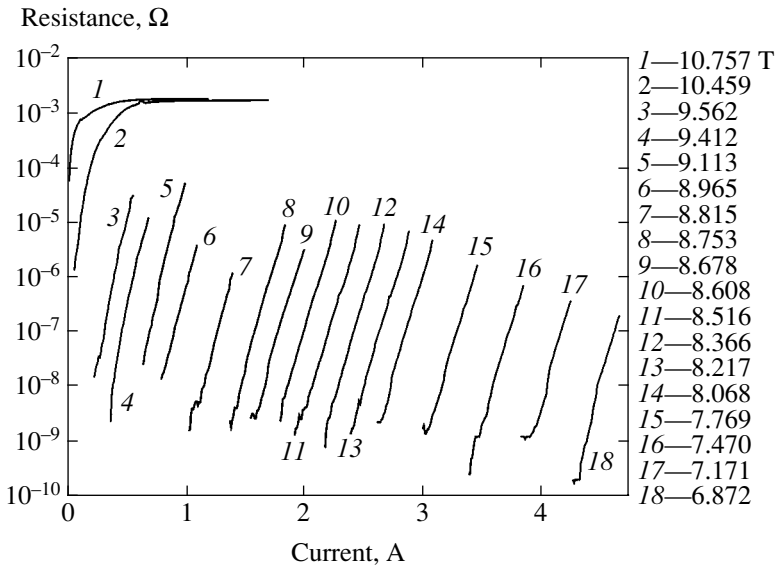


Fig. 3. Experimental resistance–current curves of sample 2 measured in various magnetic fields (all curves except *1* and *2* are corrected for the current flowing in the copper coating).

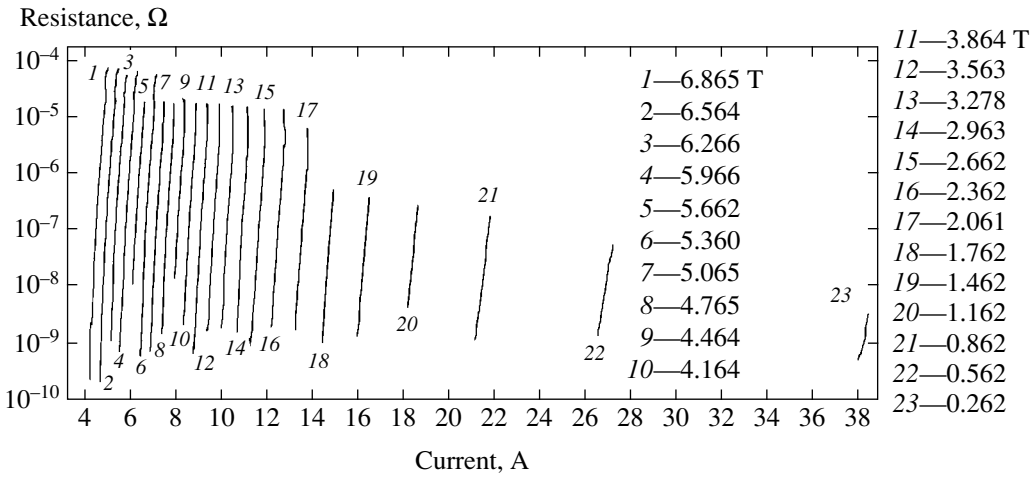


Fig. 4. Experimental resistance–current curves of sample 3 measured in various magnetic fields (all curves are corrected for the current flowing in the copper coating and stainless steel substrate).

magnetic fields, because it might readily burn out (before the action of the protecting scheme) in cases of loss of stability. In this region, the *V–I* measurements can be performed with satisfactory accuracy using sample 2 (Fig. 3). The *V–I* curves of sample 3 (Fig. 4) are stable in the entire range of fields, but the accuracy in determining the current flowing in the superconductor in high fields is limited by the uncertainty in the current (of comparable magnitude) flowing in the normal metal. For this reason, sample 3 was used for the *V–I* measurements in low fields, where the risk of burning out for sample 2 was large.

As the external magnetic field was decreased from 11.66 T to zero, the critical current increased from 1 mA to 40 A. The results of *V–I* measurements were used to calculate the resistance as a function of the current for various magnetic fields. The resistance–current curves are presented in Figs. 2–4. These dependences have proved to be exponential in a rather broad range of variation of the sample resistance, which allowed the plots (in semilogarithmic coordinates) to be extrapolated by the least-squares method to a zero current. This yielded the values of a certain conditional resistance of the superconducting core for various magnetic field at a zero current (Fig. 5). The accuracy of determining the logarithm of this resistance was better than 1% for the overwhelming majority of resistance–current curves, and the extrapolation error was always significantly smaller than the size of symbols in Fig. 5. Sample 1 was also used for detecting the field-dependent transition, that is, the dependence of the resistance on the magnetic field at a fixed probing current (1 mA).

3. DISCUSSION OF RESULTS

Measurements of the voltage–current curves performed in the fields close to the critical value make it possible to study the laws governing the current-dependent transitions at very small currents in the absence of instability. This knowledge clarifies questions concerning the adequacy of our notions about the flux flow regime in superconductors with high pinning. The traditional approach to description of the transition characteristics [27–30] was developed within the first three years after the discovery of materials possessing high current-carrying capacity, when researchers were greatly impressed at finding a rigid vortex lattice determining the properties of type II superconductors. According to these notions, the vortex dynamics in superconductors with high pinning is identical to the lattice dynamics in the ideal type II superconductor. The electric field in this resistive state (called the flux flow regime) linearly increases with the current, and the differential resistance does not exceed the normal resistance:

$$\frac{\partial E}{\partial j} = \rho_n \frac{B}{B_{c2}}. \quad (2)$$

Thus, the superconductor under isothermal conditions occurs in the resistive state and does not pass to the superconducting state at an arbitrarily high current. The possibility of reducing the properties of a superconductor with high pinning to those of the ideal type II superconductor seemed a plausible hypothesis, but this hypothesis was never experimentally confirmed for a material with sufficiently high current-carrying capacity. It was commonly accepted that the flux flow regime cannot be observed in superconductors with high critical currents because of thermal instability, which is developed already in the exponential region of the voltage–current curve [32]. Nevertheless, it was believed that this regime actually exists, in particular, in wires stabilized with normal materials. This hypothesis underlies the classical theory of stability of superconducting wires [33]. The traditional description, reproduced in numerous monographs [6, 9, 34–38] and frequently treated as the only correct approach, has become the basis of the modern, rather contradictory scalar electrodynamics of technical superconductors.

According to the inhomogeneous superconductor model, an increase in the transport current results in the critical state being attained locally, at certain low sites, and the current is redistributed in the wire cross section unless the entire sample bulk would gradually pass to the normal state. The characteristic describing this process has the shape of logistic curve (1). We failed to measure this characteristic in the entire range and only observed it in the vicinity of the critical field, for currents below 20 mA.

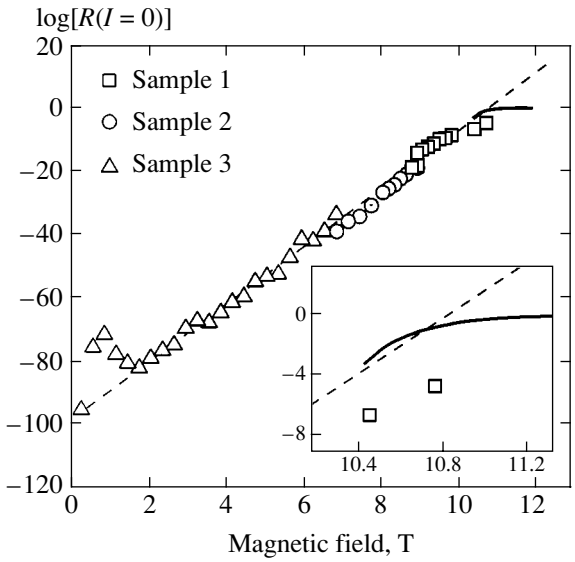


Fig. 5. Field dependence of the resistance of a niobium-titanium wire determined by extrapolation to a zero current, in comparison to the curve measured independently in sample 1 using a probing current of 1 mA.

The voltage–current curves presented in Fig. 2a contain no sufficiently extended linear portions that might be considered as manifestations of the flux flow regime. As the current increases, the voltage drop gradually reaches a value for the normal metal, while the differential resistance in the upper part of the transition significantly exceeds the normal resistance. We believe that this discrepancy with the behavior predicted by relation (2) is evidence in favor of our point of view, according to which high pinning superconductors do not feature the flux flow regime. The resistance–current curves of the material studied in this range of fields (Fig. 2b) resemble the logistic curves for the temperature- and field-induced transitions, except that the initial resistance significantly differs from zero. The existence of this initial resistance is explained by the two-component nature of the niobium–titanium alloy established in [2], whereby the body and surface (where the impurities are concentrated) of grains are characterized by somewhat different critical fields. For this reason, the current-dependent transition at the grain boundaries proceeds on the background of a pedestal corresponding to the field-dependent transition in the grain body. In Fig. 2b, triangles represent the values of sample resistance measured for the field-dependent transition at a probing current of 1 mA (for the same values of the external field as those involved in the *V–I* measurements). Naturally, the observed resistance–current curves pass through these points.

Any resistance–current curves must begin at the point where the resistance corresponds to the given magnetic field at a zero current (this resistance is usually assumed to be zero). In order to find such points

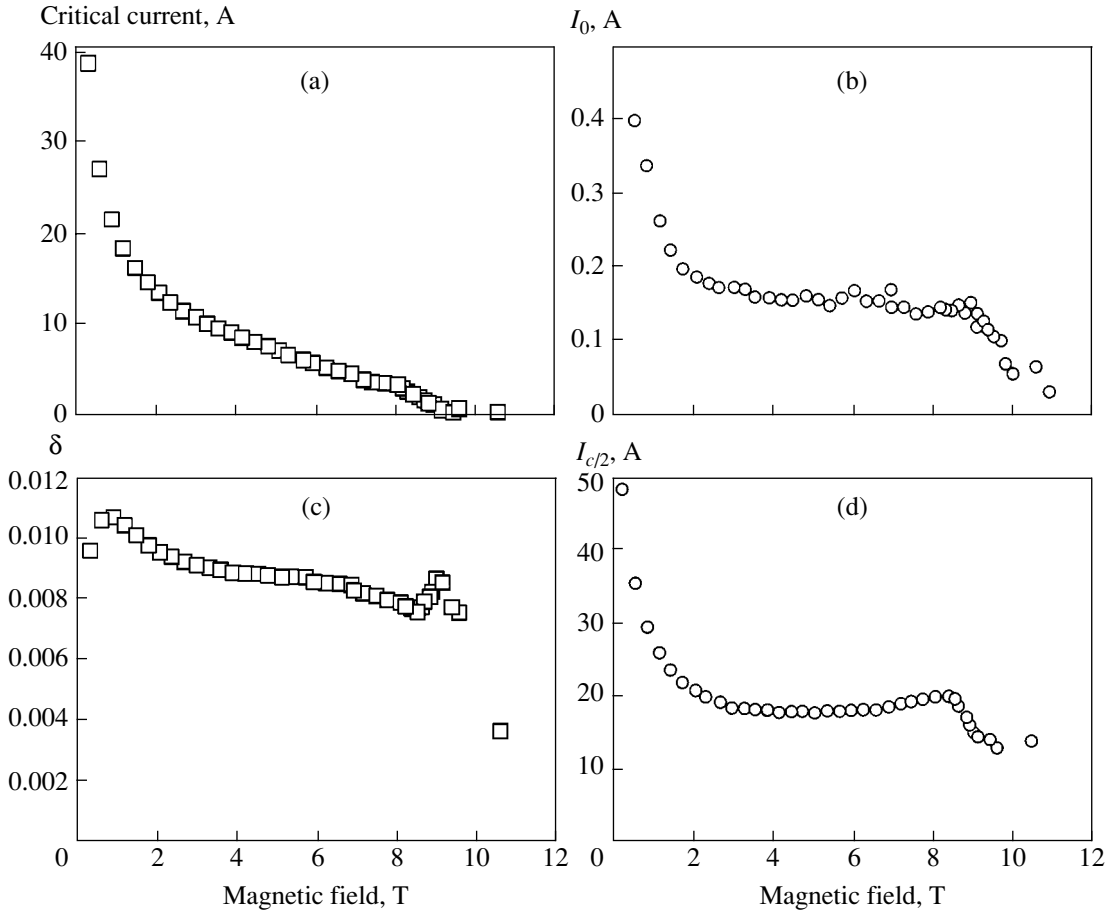


Fig. 6. Field dependences of (a) the critical current in a niobium–titanium wire (for the criterion E_c = 1 μV/cm), (b) the inverse increment I_0 of the resistance–current curve, (c) the calculated inhomogeneity parameter δ, and (d) the maximum current $I_{c/2}(0, 0)$ on a critical surface corresponding to half of the normal resistance.

corresponding to the fields significantly lower than B_{c2}, we extrapolated the exponential resistance–current curves toward the zero current (Fig. 5). It was found that the dependence of the extrapolated values on the magnetic field is well described by an exponent up to a value as small as 10^{-85} Ω. This value is physically meaningless and (within the framework of the statistical treatment) it only implies that the assembly of inhomogeneous particles comprising a superconductor contains those with critical temperatures, fields, and currents arbitrarily close to zero although the probability of such particles covering the current trajectories formed by particles occurring in the superconducting state rapidly decreases with current, magnetic field, and the temperature. In accordance with expressions (1), the dependence of this latent resistance on the field is a continuation of the resistance–field curve measured at a small probing current.

The results of the experimental data processing presented in Fig. 5 show good agreement with Eq. (1). Using the whole body of experimental data, it is possible to construct the field dependences of fitting parameters of the model under consideration, namely, of the δ value and the current density corresponding to half of the normal resistance:

$$j_{c/2} = j^*_{c/2}(T, B)\left[\left(1 - \frac{T}{T_c(0)}\right)\left(1 - \frac{B}{B_c T}\right)\right]^{-1}.$$

The calculations were performed using the values of current related to the current density as $I_{c/2}(H_e) = Sj_{c/2}(\mu_0 H_e)$. This relation, which is not as trivial as it might seem at first glance, is valid in a broad range of fields except the region of low fields where a point of zero magnetic induction appears, owing to the contribution of the current self field, in the wire cross section [39]. The parameters δ and $I_{c/2}$ were determined for each current-dependent transition using the relation following from Eq. (1):

$$\delta = \frac{(1 - T/T_c)[1 - B/B_{c2}(T)]}{[I_c/I_0 + \ln(E_c/I_c R_n)]}, \qquad (3)$$

where $I_0 = I_{c/2}\delta$, and the critical temperatures and fields correspond to half of the normal resistance (these values were determined previously [40]). The critical cur-

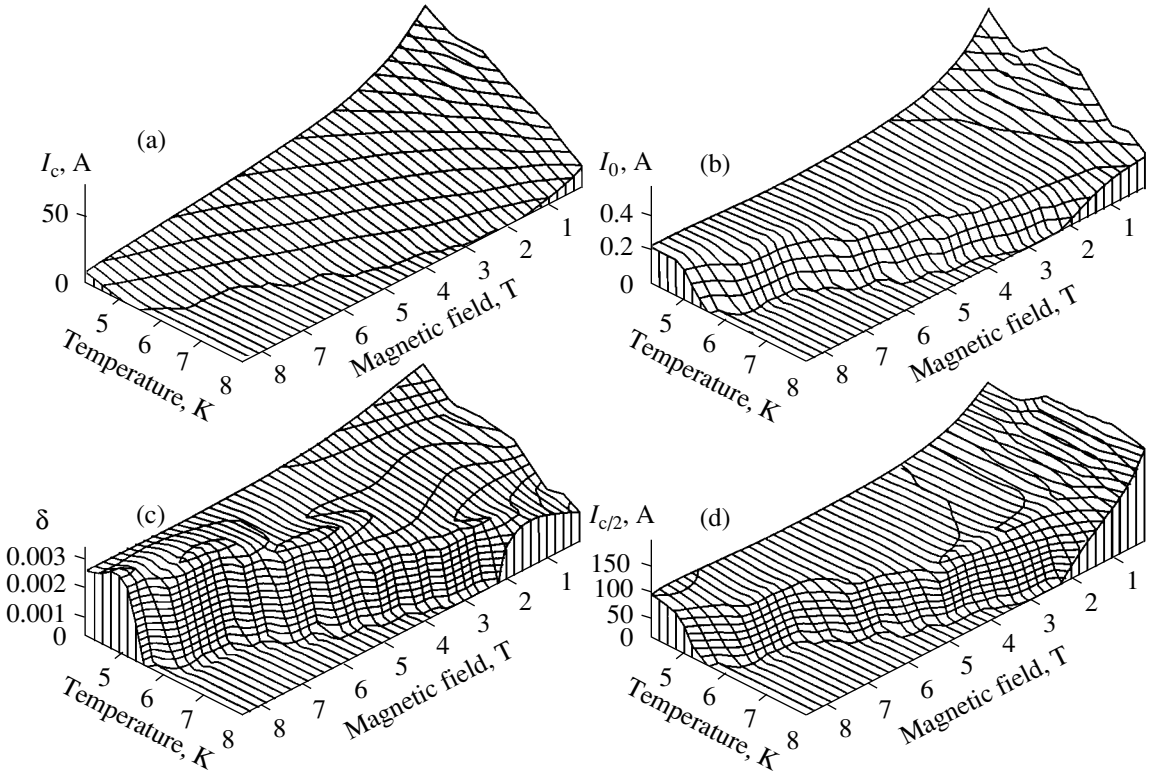


Fig. 7. Characteristics of the current-dependent transition in a niobium–titanium wire with a diameter of 120 μm as functions of the temperature and magnetic field (data from [40]): (a) the critical current (for $E_c = 1$ μV/cm); (b) the inverse increment I_0 of the voltage–current curve; (c) the relative variance of critical characteristics of the wire; (d) the calculated critical current $I_{c/2}$ corresponding to half of the normal resistance at zero temperature and induction (see relations (5)).

rent I_c was determined using the traditional criterion $E_c = 1 \times 10^{-4}$ V/m. The value of I_0, equal to the inverse current increment, was determined from the slope of the resistance–current curve.

Figure 6 presents the results of the primary processing of experimental data and the calculation of fitting parameters. It was found that these parameters change but slightly in a broad range of fields. Deviations of the $I_{c/2}$ value from a constant level in the regions of high and low fields have a quite simple explanation confirmed by independent experiments [2, 3]. As the field increases above 9 T, inhomogeneity of a real material is manifested and some grains already exhibit a transition to the normal state in this field. The observed decay in I_0 and $I_{c/2}$ corresponds to contraction of the effective superconductor cross section caused by the appearance of normal bulks in the wire. In the fields below 1.5 T, anisotropy of the wire material becomes significant. Indeed, it was demonstrated [3] that thermomechanical treatment of a wire results in the formation of a texture characterized by cylindrical symmetry. The pinning in such a wire is characterized by a tensor [1, 3] having three main components along the axis and in the radial and azimuthal directions in the transverse cross section. The pinning is much higher for vortices in the azimuthal direction than for those in the radial direction. In moderate and high fields, where the direction of magnetic induction in the material coincides with the external field, the pinning tensor exhibits averaging. In a low field, the presence of the current self field makes the azimuthal component prevailing over the radial one. This gives rise to a sharp growth in the extent of pinning and, hence, in the $I_{c/2}$ and I_0 values.

The results of processing of the experimental data described above made it necessary to return to our early experiments carried out in 1978, where the voltage–current curves of niobium–titanium and niobium–zirconium wires were studied (of course, using worse measuring equipment) as dependent both on the magnetic field and on the temperature. The experimental procedure was described in detail and the primary results were tabulated in [40]. Application of the method of data processing described above showed (Figs. 7 and 8) that the characteristic constants δ and $j_{c/2}$ retain their values not only in a wide interval of fields, but in a broad range of temperatures as well.

The results drive us to revise the commonly accepted approach to description of the current-carrying capacity of superconductors with high pinning. It has been established that the resistive state of such a superconductor (at least, for Nb–50 wt % Ti and Nb–50 wt % Zr solid solutions studied) can be

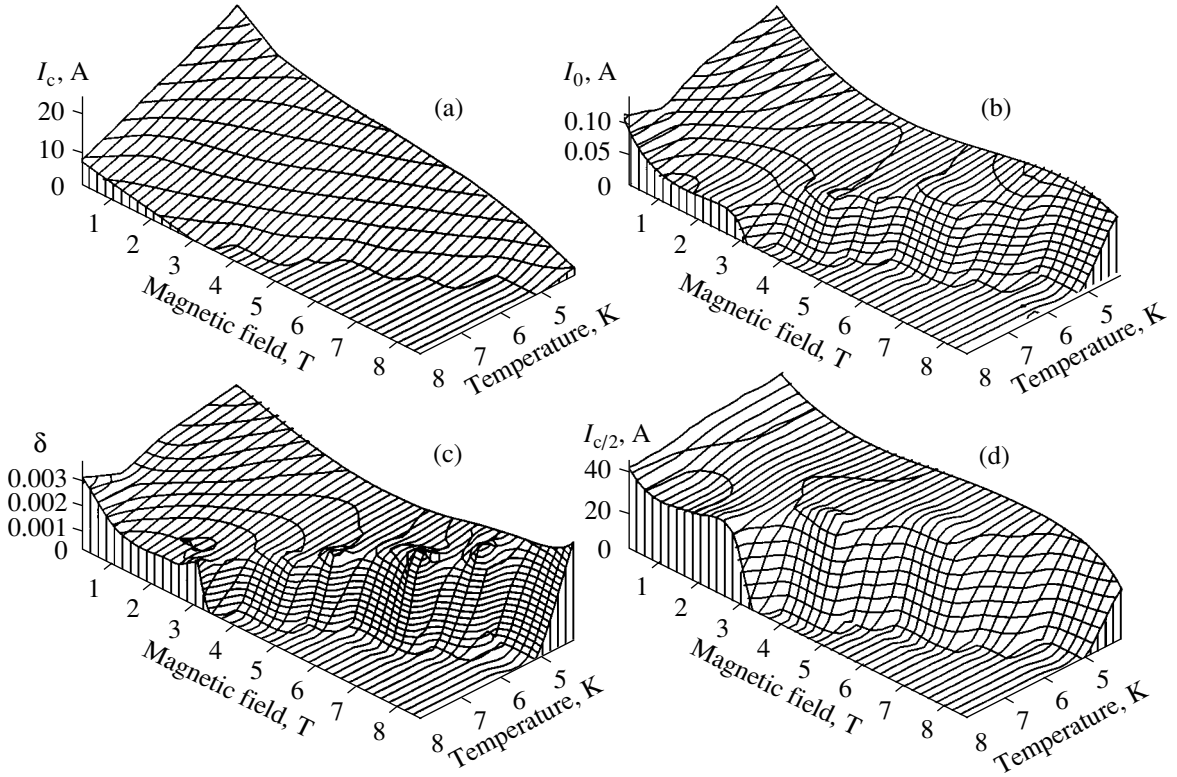


Fig. 8. Characteristics of the current-dependent transition in a niobium–zirconium wire with a diameter of 80 μm as functions of the temperature and magnetic field (data from [40]): (a) the critical current (for E_c = 1 μV/cm); (b) the inverse increment I_0 of the voltage–current curve; (c) the relative variance of critical characteristics of the wire; (d) the calculated critical current $I_{c/2}$ corresponding to half of the normal resistance at zero temperature and induction (see relations (5)).

described in terms of the $B_{c2}(T)$ function and three constants: $T_c(0)$, δ, and $j_{c/2}$. If the real anisotropy of a superconductor can be ignored, the last constant is the current density corresponding to half of the normal resistance of the given material at a zero temperature in a zero magnetic field (if the attainment of this point were not hindered by thermal instability). The conditional quantity, usually called critical current, which corresponds to the state in which the superconductor resistance is R_c, is determined from Eq. (1) as

$$I_c = I_{c/2}\left(\delta \ln\frac{R_c}{R_n} + \left(1-\frac{T}{T_c}\right)\left(1-\frac{B}{B_c(T)}\right)\right). \qquad (4)$$

Ignoring the small term $\delta\ln(R_c/R_n)$, we obtain the well-known scaling

$$F_{Lc} = F_{Lcmax} b(1-b),$$

where

$$b = \frac{B}{B_{c2}(T)},$$

$$F_{Lcmax}(T) = B_c(T)\left(1-\frac{T}{T_c}\right) I_{c/2},$$

and F_{Lc} is the critical Lorentz force (we avoid denoting the last quantity as F_p and using the term "pinning force," since this may lead to incorrect treatment of this value as a vector [1]). The dome shape of $F_{Lc}(b)$, which is usually attributed to structure-sensitive pinning, in our case is related to the field dependence of the resistance at the starting point of the resistance–current curve

$$\rho_{eff} = \rho_n \exp\left[-\left(1-\frac{T}{T_c(0)}\right)\left(1-\frac{B}{B_{c2}(T)}\right)/\delta\right],$$

involving no parameters directly depending on the pinning structure. Using a scaling of the type

$$F_{Lc} = F_{Lcmax} b^{\mu}(1-b)^{\nu},$$

it is necessary to consider the fact that pinning accounts for only a fraction of this dependence, namely, $b^{\mu-1}(1-b)^{\nu-1}$. Refraining here from speculating on possible factors responsible for the deviation of exponents from unity in other superconductors, we only emphasize that many technical superconductors are in fact mixtures of several superconductors [23].

An important feature of Eq. (1), which was not considered before, is the fact that the width of the superconducting transition with respect to the temperature, field, and current is controlled by the same parameter δ. It was claimed above that, irrespective of the nature of a particular inhomogeneity, the temperature-, field-, and current-dependent transitions exhibit simultaneous and proportional broadening. The basic reasons underlying this relationship can be understood using the notion of the critical layer (Fig. 9) [14], which represents a generalization of the widely used concept of the critical surface. The critical surface is the set of points separating the normal and superconducting states in the $\{T, B, j\}$ coordinate space, under the assumption that the transitions exhibit a jumplike character. In materials with smoothed superconducting transition, the critical current is a conditional quantity corresponding to a certain conventionally selected criterion, for example, an effective resistivity equal to 10^{-12} Ω m. This criterion can be represented by a "critical surface" of its own. Selecting another criterion, we obtain a different critical surface (not intersecting with the first one). Thus, the superconducting and normal states in the $\{T, B, j\}$ space are separated by a transition layer, each point of which corresponds to a certain effective resistance, and the points with equal effective resistance form nonintersecting critical surfaces. Displacements along each coordinate between the neighboring critical surfaces are related as

$$\Delta\rho_{\text{eff}} = \frac{\partial\rho_{\text{eff}}}{\partial T}\Delta T = \frac{\partial\rho_{\text{eff}}}{\partial B}\Delta B = \frac{\partial\rho_{\text{eff}}}{\partial j}\Delta j. \qquad (5)$$

These relations clearly demonstrate that the superconducting transition is simultaneously smoothed with respect to the temperature, field, and current with the same relative variance. Can statistical broadening of the transition be eliminated? A system in which the physical factors influencing broadening of the current-dependent transition can be studied is probably offered by a single crystal containing homogeneously distributed identical pinning centers and featuring sufficiently steep field- and temperature-dependent transitions. However, if the depth of the potential well of introduced centers or their density acquire a finite variance, the temperature- and field-dependent transitions will be smoothed to the same degree and the initially homogeneous crystal will become the classical inhomogeneous material.

We believe that the results presented above give us grounds to use the hypothesis about inhomogeneity of real superconductors in the development of vector electrodynamics, with a view to its possible verification by means of comparison with the results of a large number of recent experimental investigations. At the same time, the above considerations show that this hypothesis implies revision of some canonical concepts and methods of analysis, so that rather serious competition with the existing notions can be expected. However, we believe that the competition is possible with another hypothesis, rather than with a system of well-established facts. For this reason, we consider it expedient to provide this paper with an Appendix presenting our understanding of the present-day state of the activation model.

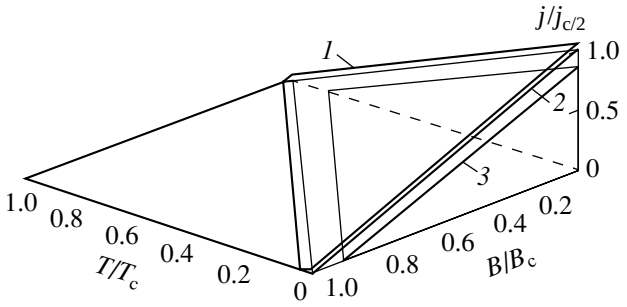


Fig. 9. Schematic diagram illustrating the notion of a critical layer. Critical surfaces *1–3*, corresponding to the effective resistances amounting to 0.9, 0.5, and 10^{-6} of the normal resistance at $\delta = 0.01$, respectively, are parallel to each other. The position of line $j = 0$ depends on the selected level of effective resistance.

4. CONCLUSIONS

A comparison of the results of experimental investigation of the transition characteristics of a niobium–titanium wire to the inferences of a phenomenological model of an inhomogeneous superconductor with high pinning showed good agreement. The generalized superconducting transition is well described in terms of the exponential dependence of the effective resistance of a superconductor on the temperature, magnetic induction, and current density. Only two fitting parameters, δ and $j_{c/2}$, are sufficient in order to describe the temperature-, field-, and current-dependent transitions in a broad range of magnetic fields and temperatures.

It is necessary to take into account the tensor character of pinning for correct description of the generalized superconducting transition in low magnetic fields, where the current self field significantly influences the direction of the total induction. Near the critical magnetic field, the transition characteristic is affected when the effective cross section of the superconducting wire is reduced caused by the transition of a certain fraction of grains into the normal state in the fields significantly lower than the B_{c2} value determined at a small probing current. Retention of the trace resistance at arbitrarily low values of the temperature, induction, and a zero current in Eq. (1) is related only to the inhomogeneity of a superconducting material.

We are not inclined to believe that our results reject the Anderson model, but it is obvious that we have to find a way to distinguish the effects of thermal activation on the background of another, statistical nature of broadening of the voltage–current curves, so as to

obtain reasonable estimates of the parameters of individual pinning centers.

APPENDIX

Models of Thermally Activated Magnetic Flux Creep and Quantum Tunneling

The model of thermally activated vortex motion was developed by Anderson [5] in 1962 upon analysis of the experimental data of Kim *et al.* [27] on the logarithmic decay of magnetic field in a niobium–zirconium cylinder and the exponential voltage–current curves of a niobium–zirconium wire. If the experimenters could have noticed that their resistance–current curves were exponential as well, the model would probably be different. However, Anderson suggested that the motion of vortices in a superconductor is analogous to the motion of dislocations in a stressed metal [41], which is manifested by a temperature-dependent creep of metals. Based on this analogy, Anderson obtained the well-known expression for the voltage–current curve of a superconductor:

$$E \propto \exp\left(\frac{U_0 - pjB}{k_B T}\right). \quad \text{(A.1)}$$

The notion about thermally activated vortex motion in a medium with potential relief was attractive because it gave hope of explaining the observed linear temperature dependence of the critical current. For a conditionally selected critical value of the electric field E_c, the critical current is described by the relation

$$j_c = \frac{1}{pB}\left(U_0 - kT\ln\frac{E_c}{E_0}\right), \quad \text{(A.2)}$$

which reflects a decrease in the critical current with increasing temperature and magnetic field in agreement with the behavior observed in experiment. In addition, this theory explains logarithmic decay of the magnetic moment of a magnetized superconductor and the field dependence of the critical current (previously proposed by Kim *et al.* [27]),

$$j_c = \alpha(T)/(H + B_0), \quad \text{(A.3)}$$

where the term B_0 was introduced in order to eliminate divergence of the critical current density when the external field decreases to zero. According to Anderson [5], $B_0 = \Phi_B/d^2$, where Φ_B is the magnetic flux of a vortex bundle and d is the bundle diameter.

The model of thermal activation was the first theory capable of describing the results of the large number of experiments carried out by that time. This theory offered an elegant and clear physical picture of the motion of magnetic flux in a superconductor and it conformed well to then adopted "physical fashion," which explains why it was so popular and included into all handbooks and monographs. Nevertheless, the model met a certain criticism from the very beginning, and the experiments revealed serious discrepancies—such that would be fatal for any other model. However, the lifetime of the thermal activation model was prolonged at the expense of admission of an arbitrary dependence of the model parameters on the temperature, field, and current. This marked the origin of a very fruitful branch in the theory of pinning, which related these dependences to various microscopic features of the interaction between vortices and trapping potential wells and made it possible to explain any experimental result within the framework of the thermal activation model. We believe that it was a remarkable fact that Anderson in a later review [42] made a reference to his most frequently cited paper [5] only in the context of a discussion about vortex motion under the action of the Lorentz force, not even mentioning the activation model. In our opinion, it was the absence of rather active discussions (in 1963–70s) in the mentioned monographs that accounted for a revival of the thermal activation model upon the discovery of high-temperature superconductivity. The revival was inspired by the paper of Dew-Hughes [43], where it was claimed that, if the thermal activation model is valid, its predictions should be especially clearly manifested in HTSC at elevated temperatures. According to these predictions, new HTSC must exhibit a limited current-carrying capacity at high temperatures. However, no new evidence in favor of validity of the thermal activation model was presented.

This Appendix summarizes arguments against the thermal activation model, including both reasoning used by opponents in the early discussion mentioned above and some new arguments arising in recent years.

(i) Expression (A.1) introduces a certain discomfort by predicting a finite electric field at a zero current density. Prolonged discussion about this paradox was calmed by the discovery that such differential characteristic does not lead to a discrepancy in the integral characteristic: when the current I flows in a plate in the direction perpendicular to the external field B_e (applied in the plane of the plate), the voltage drop V becomes zero when the current ceases. The contributions due to the intrinsic magnetic field of the current have opposite signs on opposite sides of the plate, and the integral characteristic acquires the form

$$V \propto \exp(\beta B_e)\sinh(\alpha I),$$

where factors α and β make the arguments dimensionless. However, the above simple considerations by no means provide a ground to change the differential char-

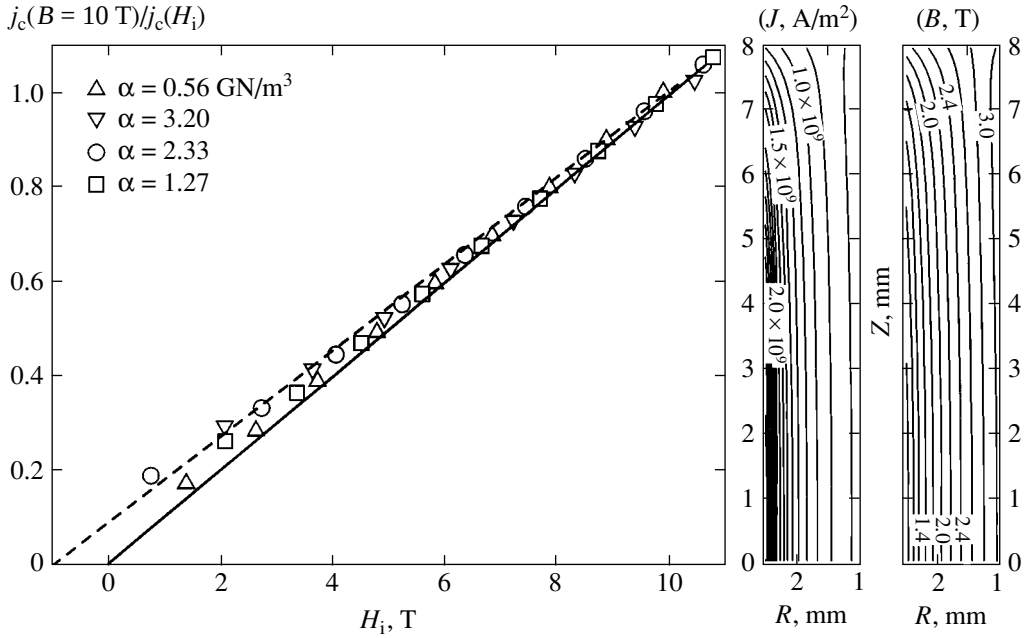


Fig. A1. Influence of a finite length of a tubular sample and self-consistent distributions of the induction and current density in the tube wall on the shape of the measured integral characteristic $J_c(H_i)$ (dashed line) for a differential characteristic of the $j_c(B) = \alpha/B$ type (solid line). The right-hand insets show the maps of self-consistent distributions of the current density J and induction B in the tube wall.

acteristic so as to render it a form eliminating the discussion [6]:

$$E = \nu_0 Bl \exp\left(-\frac{U_0(T, B)}{k_B T}\right)\sinh\left(\frac{U_1(T, B)}{k_B T}\right). \quad (A.4)$$

Confusing the integral and differential characteristics is a permanent source of ambiguity in the theory of pinning.

(ii) Let us consider expression (A.4), where U_0 is the height of the pinning-related potential barrier hindering the motion of a separate vortex bundle, $U_1 = jBV_c r_p$ is a decrease in this height due to the Lorentz force, and V_c and r_p are rather ambiguously defined dimensional factors. The first of these quantities, V_c, is given a meaning of the bulk of the vortex bundle capable of surmounting the barrier independently of the other bundles; the second factor, r_p, has the meaning of a characteristic length of variation of the pinning potential. The preexponential factor includes a product of ν_0, the frequency of bundle oscillations at the pinning center, and l, the length by which the bundle displaces upon a single jump. It was admitted that any of the model parameters may depend on the magnetic field and the temperature, and that the barrier (reduced by the current) can be a nonlinear function of the current [30, 44]. Moreover, it was even assumed [17] that the hyperbolic sine argument has a logarithmic character, which was necessary to make the model capable of explaining the power shape of the voltage–current curves of superconductors. Such a number of assumptions and fitting parameters generally removes the question of experimental verification of the model. Nevertheless, many researchers are still engaged in experimental and theoretical investigations of some parameters of the thermal activation model (see. e.g., [44]).

(iii) Admitting the dependence of U_0 and U_1 on temperature, the theory refused its main original destination: explaining the linear temperature variation of the critical current. Of course, this explanation was questionable from the very beginning, since the temperature at which the critical current vanishes in Eq. (A.2) is not related to the critical temperature.

(iv) There is a quite steady opinion that the thermal activation model is "reliably confirmed" by the observed logarithmic law of decay of the magnetic moment of a superconductor. It should be recalled that experiment cannot confirm a theory: in the best case, there are merely no discrepancies between them. A simple electrotechnical problem in calculating the magnetic flux diffusion in the sample with a given voltage–current curve was repeatedly described in the literature [5, 30, 46, 47]. In order to obtain the logarithmic variation of the magnetic moment with time, it is sufficient that the sample resistance would rapidly decrease with the current density: this behavior was observed in materials having both exponential and power voltage–current curves.

It is commonly accepted that experiments of this type offer certain advantages in comparison to direct measurements of the voltage–current curve. The first

apparent advantage is that a difference in levels of the electric field makes it possible in principle to refine the shape of the *V–I* curve with a sensitivity inaccessible in the case of direct measurements. The second such advantage is apparent simplicity of taking into account the current self field as a result of measurement of the magnetic field inside a tubular sample or the magnetic moment in the magnetic field parallel to the axis of a cylindrical sample. The latter advantage is used for determining the value of B_0 in Eq. (A.3). Indeed, according to Eq. (A.3), the inverse critical current density is a linear function of the magnetic field, and B_0 is given by the intercept of this line with the field axis. The linear current density in experiments with tubular samples is determined from the difference in magnetic fields outside and inside the tube [5, 45], assuming that the sample length is sufficient for the linear relationship to be valid. However, it is this assumption that leads to a regrettable error. This is illustrated in Fig. A1, which shows the results of simulation of the typical experiment. Calculations were performed for a sufficiently complete set of data taken from [45], where the properties of molybdenum sulfides were studied using tubular samples with a length of 16 mm, an outer diameter of 5 mm, and a wall thickness of 2 mm. Using the method described in [1], we obtained self-consistent distributions of the current density and magnetic induction in the wall for four samples with the known maximum pinning force. It is important to note that we used a differential characteristic of the type corresponding to Eq. (A.3) and assumed that $B_0 = 0$. Then, the inverse difference of magnetic fields outside and inside the tube was plotted as a function of the magnetic field. These plots were found to be linear, but their slopes were lower than the slope of the differential characteristic and the points of intersection with the field axis were situated to the left from zero. The scale of the effect coincided with the results of determination of B_0 presented in [45, Fig. 4]. This situation is analogous to that reported previously [39], where it was demonstrated that the behavior of $I_c(H_e)$ in the region of low fields is determined by the current self field, an that a singularity at the zero of the differential characteristic $j_c(B) = \alpha/B$ does not lead to the appearance of singularities in the integral characteristic in the zero external field. Thus, the very introduction of the parameter B_0, is a consequence of erroneously mixing up the integral and differential characteristics.

(v) In order to bring the temperature dependence of the slope of the voltage–current curve into agreement with experiment [10], it is necessary to admit that $V_c r_p \sim T/(1 - T/T_c)$. However, this assumption is obviously strains the point too much.

(vi) In 1972, Campbell and Yvetts [9] pointed out that the only advantage of the thermal activation model, in their opinion, is that it predicts the exponential shape of the voltage–current curve. However, it was found that, on the one hand, nonexponential voltage–current curves can also conform to the activation scheme [17] and, on the other hand, the exponential shape of the initial portion of any superconducting transition can be explained in terms of a bulk inhomogeneity of the superconductor [15].

(vii) Although the thermal activation model has proved to be highly robust with respect to criticism, the superconductor community sometimes recognizes limited applicability of this theory. One example is offered by the quantum tunneling model [45]. The rate of field decay inside a tubular sample upon switching off the external field is defined as

$$\frac{1}{H_i}\frac{dH_i}{d\ln(t/t_0)}.$$

According to the canonical Anderson model [5], this rate is expressed as

$$-\frac{4\pi w}{H_i(H_i + B_0)}\frac{k_B T}{d^4}.$$

It was found that, as the temperature decreases, the field decay rate tends to a finite nonzero value. If the decay were caused by thermal fluctuations, it would not take place at zero temperature, which makes the hypothesis of the quantum tunneling of vortices quite logical. However, if we take into account that $\Phi_B d^2$ in [5] corresponds to the product $V_c r_p$ in Eq. (A.4) and that the latter product also tends to zero with decreasing temperature, this hypothesis does not seem as obvious.

Let us consider another aspect: inhomogeneity of a material is retained at zero temperature and this reason for the field decay may compete with the hypothesis of quantum diffusion. Considering possible factors responsible for the decay, Mitin [45] rejected the model of a bulk-inhomogeneous superconductor because he believed that this model, first, necessarily implies a linear relation between the critical current and the field and, second, it predicts the decay rate as determined by the relative degree of inhomogeneity of the material. According to the experimental results obtained in [45], the relation between the critical current and the field was nonlinear, while the decay rate was lower in a highly inhomogeneous sample than in a more homogeneous one. However, the first reason is erroneous because inhomogeneity cannot change the relationship between critical current and field (which is determined by physical, rather than statistical factors). In our paper [14] cited in [45], a linear critical surface was selected as the simplest case in order to demonstrate a relationship between the broadening of transitions with respect to the current, field, and temperature (the linear behavior of $j_c(H)$ observed in this article needs physical explanation and is probably not universal). The second reason is also erroneous, since even the rough expression (written without any justification) shows that the decay

rate is determined by a product of the relative variance and the critical current, while the critical current for the inhomogeneous sample was about fifty times as small as that of the homogeneous one.

In order to finally eliminate the latter confusion, we will obtain an expression for the rate of field decay inside a thick tube made of an inhomogeneous superconductor with a linear critical surface. This derivation generally reproduces a procedure used to obtain an analogous relation in [30], where it was originally established that the problem of description of the electric field diffusion is the simplest to solve. It is especially interesting to reproduce the calculation, since the result (coinciding with experiment) is obtained without recourse to the thermal activation model. In order to simplify expressions, let us solve the problem in a dimensionless form for a flux frozen between two parallel plates. From the standpoint of physics, this problem is identical to the description of experiments with superconducting tubes. Consider two plates of thickness d, which are parallel to the xz plane, and let the field be oriented along the z axis. The screening currents and electric fields in one plate are directed along the x axis, while those in the other plate are oriented in the opposite direction. Taking into account that all parameters depend only on the time and the y coordinate, we formulate the quasi-stationary problem as

$$\begin{gathered}\operatorname{curl} b = i,\\ \frac{\partial \varepsilon}{\partial \xi} = \frac{\partial b}{\partial \tau},\\ \varepsilon = i\exp\frac{-1+\theta+|b|+|i|}{\delta},\end{gathered} \tag{A.5}$$

where $b = B/B_{c2}$, $\theta = T/T_c$, $i = j/j_{c2}$, $\varepsilon = E/\rho_n j_{c/2}$, $\xi = \mu_0 j_{c/2} y/B_{c2}$, and $\tau = \rho_n j_{c/2} t/B_{c2}$.

In order to specify the initial and boundary conditions, let us assume that macroscopic currents in the plates before the experiment are absent, and that a homogeneous and sufficiently high magnetic field b_e exists in a part of the space containing the plates. After switching off the field source, the rapid process of magnetic flux redistribution involves the whole plate thickness, so that the field b_i frozen between the plates becomes smaller than the initial value (b_e). Considering one of the plates, we can write the boundary conditions as $b(0, \tau) = b_i(\tau)$, $b(\xi_0, \tau) = 0$ (ξ_0 is the dimensionless thickness of the plate). Following the method described in [30], the electric field diffusion in the plate is described by a separable differential equation

$$\delta\frac{\partial \ln(\varepsilon)}{\partial \tau} = -\frac{\partial \varepsilon}{\partial \xi} - \frac{\partial^2 \varepsilon}{\partial \xi^2}. \tag{A.6}$$

In order to obtain this expression, we differentiated the material equation (third equation in (A.5)) with respect to time and replaced the derivatives of the field and current by their derivatives with respect to the coordinate using the first two equations in (A.5). We also took into account that the field and current decrease in magnitude, and we ignored the term $\partial(\ln i)/\partial\tau$ in comparison to $\delta^{-1}(\partial|b|/\partial\tau + \partial|i|/\partial\tau)$, which is justified for not too small currents (i.e., for not very long times). Note that neglecting $\partial(\ln i)/\partial\tau$ for large times leads to a physically senseless result, whereby the fields and currents acquire negative values of arbitrarily large magnitude, rather than tending to zero. This paradox, arising in the exact solution of the problem of diffusion with a differential characteristic (A.1) according to the thermal activation model, is directly related to finiteness of the electric field at zero current density. However, neglecting $\partial(\ln i)/\partial\tau$ for reasonable experimental times is quite correct.

The solution to Eq. (A.6) has the form

$$\varepsilon = \frac{\delta}{\lambda(\tau-\tau_0)}[C_2 - \exp(C_1 - \xi) - \lambda\xi], \tag{A.7}$$

where λ is an arbitrary constant appearing due to the separation of variables (the parameter C_2 is not necessary and will be taken equal to zero). The origin of time is conveniently selected as $\tau_0 - 1$. Using this solution, we can write an equation for the magnetic field (the second equation in (A.5)) for the given boundary conditions:

$$\frac{\delta(\exp(C_1-\xi)-\lambda)}{\lambda(\tau+1)} = -\frac{\partial|b|}{\partial\tau}. \tag{A.8}$$

This equation yields

$$|b| = C_3(\xi) - \frac{\delta}{\lambda}[\exp(C_1-\xi)-\lambda]\ln(\tau+1), \tag{A.9}$$

where the absence of an external field corresponds to $C_3(\xi_0) = 0$ and $\lambda = \exp(C_1 - \xi_0)$. The value of $C_3(0)$ can be approximately determined from the condition that the plate occurs in the critical state, whereby the exponent in the material equation is zero:

$$i = \frac{\partial b}{\partial \xi} = 1 - \theta - b(\xi), \tag{A.10}$$

$$C_3(0) = (1-\theta)(1-\exp(-\xi_0)), \tag{A.11}$$

$$\begin{aligned} b_i(\tau) &= [1-\exp(-\xi_0)]\\ &\times[1-\theta-\delta\exp(\xi_0)\ln(\tau+1)].\end{aligned} \tag{A.12}$$

The argument of the logarithmic term at the initial moment is nonzero, which eliminates singularity at this point. It should be noted that an analogous expression describes the process of establishing the critical state

in [46]. In contrast, we solve the problem for an established critical state of the system. Therefore, we have no grounds to relate this process to an initially low (on the semilogarithmic scale) rate of the field variation inside the tube.

For a sufficiently large time ($\tau \gg 1$), the experimentally measured rate of field decay between the plates (normalized to the initial value) in the dimensionless form is as follows:

$$\nu(T) = \frac{1}{B_i(0)}\frac{\partial B_i}{\partial \ln(t)} = \frac{\delta \exp(\mu_0 d j_{c/2}/B_{c2})}{1 - T/T_c}. \quad \text{(A.13)}$$

This expression qualitatively describes the results reported in [45], including the finite rate of field decay at zero temperature and its linear growth at $T \ll T_c$.

The above result shows that the hypothesis of quantum tunneling is insufficiently justified, but we are more interested in consequences related to the thermal activation model. Since, as was noted above, public opinion allows the model constants to be treated as functions of the parameters of state, one can readily choose the temperature dependence such that it will compensate the growth in the absolute value of the exponent when the temperature approaches zero. The fact that $\nu(0)$ differs from zero refutes the thermal activation model no more significantly than do many other facts. However, it is hoped that the logarithmic field decay law obtained for a voltage–current curve different from (A.1) must violate the widespread opinion that this law unambiguously confirms the validity of the thermal activation model. Apparently, only further accumulation of alternative explanations of the results of experimental investigations, now traditionally interpreted in terms of this rather inconvenient model, will eventually reject this theory. The statistical model of transition characteristics is free of obvious discrepancies of this kind but, unfortunately, it closes the channel for obtaining data about the microscopic interactions between vortices and the point pinning centers.

We can hardly expect that the radical point of view, according to which the apparently elegant Anderson hypothesis was unrealistic from the very beginning, will be quickly accepted. However, there are no reasonable arguments against the other approach to interpretation of the existing situation. Of course, the process of thermal activation of the vortex motion can be important. However, the bulk inhomogeneity of superconductors with high pinning has been reliably established and unambiguously influences the shape of the transition characteristics. Therefore, investigation of the activation processes requires the development of an approach capable of distinguishing their contribution on the background of a rough statistical part. Unfortunately, to the best of our knowledge, no such attempts have been reported.

REFERENCES

1. E. Yu. Klimenko, S. V. Shavkin, and P. V. Volkov, Zh. Éksp. Teor. Fiz. **112**, 1055 (1997) [JETP **85**, 573 (1997)].
2. E. Yu. Klimenko, S. V. Shavkin, and P. V. Volkov, Fiz. Met. Metalloved. **92** (6), 27 (2001) [Phys. Met. Metallogr. **92**, 552 (2001)].
3. E. Yu. Klimenko, M. S. Novikov, and A. N. Dolgushin, Fiz. Met. Metalloved. **92** (3), 11 (2001) [Phys. Met. Metallogr. **92**, 219 (2001)].
4. A. Gurevich and V. M. Vinokur, Phys. Rev. Lett. **83**, 3037 (1999).
5. P. W. Anderson, Phys. Rev. Lett. **9**, 309 (1962).
6. J. B. Ketterson and S. N. Song, *Superconductivity* (Cambridge Univ. Press, Cambridge, 1994).
7. W. J. Carr, Jr., *AC Loss and Macroscopic Theory of Superconductors* (Gordon and Breach, New York, 1983).
8. J. Baixeras and G. Fournet, J. Phys. Chem. Solids **28**, 1541 (1967).
9. A. M. Campbell and J. E. Yvetts, *Critical Currents in Superconductors* (Taylor and Francis, London, 1972; Mir, Moscow, 1975).
10. J. M. A. Wade, Philos. Mag. **24**, 339 (1971).
11. P. Jarvis and J. G. Park, Physica (Amsterdam) **55**, 386 (1971).
12. W. H. Warnes and D. C. Larbalestier, Cryogenics **26**, 643 (1986).
13. S. Takacs, Cryogenics **28**, 374 (1988).
14. G. L. Dorofeev, A. B. Imenitov, and E. Yu. Klimenko, Cryogenics **20**, 307 (1980).
15. E. Yu. Klimenko and A. E. Trenin, Cryogenics **23**, 527 (1983).
16. F. Voelker, Part. Accel. **1**, 205 (1970).
17. E. Zeldov, N. M. Amer, G. Koren, *et al.*, Appl. Phys. Lett. **56**, 680 (1990).
18. E. Yu. Klimenko and A. E. Trenin, Cryogenics **25**, 27 (1985).
19. Yu. V. Bobkov, N. G. Vasil'ev, S. I. Zastrozhnov, *et al.*, Fiz. Met. Metalloved. **59**, 756 (1985).
20. A. E. Pashitskii, A. Gurevich, A. A. Polyanskii, *et al.*, Science **275**, 367 (1997).
21. J. Albrecht, Ch. Jooss, R. Warthmann, *et al.*, Phys. Rev. B **57**, 10332 (1998).
22. B. Pietrass, Phys. Status Solidi A **60**, 441 (1980).
23. N. Kozlenkova, A. Shikov, V. Pantsyrnyi, *et al.*, Physica C (Amsterdam) **401**, 87 (2004).
24. I. M. Dmitrenko, N. A. Fogel', and A. G. Sidorenko, Fiz. Nizk. Temp. **8**, 1153 (1982) [Sov. J. Low Temp. Phys. **8**, 583 (1982)].
25. G. L. Dorofeev, A. B. Imenitov, E. Yu. Klimenko, *et al.*, in *Proceedings of 2nd All-Union Conference on Technical Use of Superconductivity* (Leningrad, 1984), Vol. 2, p. 51.
26. K. Hense, V. Vuller, R. Lackner, *et al.*, Physica C (Amsterdam) **401**, 214 (2004).
27. Y. B. Kim, C. F. Hempstead, and A. Strnad, Phys. Rev. Lett. **9**, 306 (1962).

28. P. W. Anderson and Y. B. Kim, Rev. Mod. Phys. **36**, 39 (1964).
29. Y. B. Kim, C. F. Hempstead, and A. Strnad, Rev. Mod. Phys. **36**, 43 (1964).
30. M. R. Beasley, R. Labusch, and W. W. Webb, Phys. Rev. **181**, 682 (1969).
31. E. Yu. Klimenko, N. N. Martovetskiĭ, and S. I. Novikov, Dokl. Akad. Nauk SSSR **261**, 1350 (1981) [Sov. Phys. Dokl. **26**, 1180 (1981)].
32. M. Polak, I. Hlasnik, and L. Krempassky, Cryogenics **13**, 702 (1973).
33. R. G. Mints and A. L. Rakhmanov, *Instabilities in Superconductors* (Nauka, Moscow, 1984) [in Russian].
34. P. G. de Gennes, *Superconductivity of Metals and Alloys* (Benjamin, Amsterdam, 1966; Mir, Moscow, 1968).
35. M. Tinkham, *Introduction to Superconductivity* (McGraw-Hill, New York, 1975; Atomizdat, Moscow, 1980).
36. Y. B. Kim and M. J. Stephen, in *Superconductivity*, Ed. by R. D. Parks (Marcel Dekker, New York, 1969), Vol. 2, p. 1107.
37. E. W. Collings, *Applied Superconductivity, Metallurgy, and Physics of Titanium Alloys* (Plenum, New York, 1989), Vol. 1.
38. V. V. Shmidt, *Introduction to the Physics of Superconductors* (Nauka, Moscow, 1982) [in Russian].
39. E. Yu. Klimenko and V. G. Kon, in *Superconductivity: Proceedings of Conference on Technical Applications of Superconductivity, Alushta-75* (Atomizdat, 1977), Vol. 4, p. 114.
40. G. L. Dorofeev, A. B. Imenitov, and E. Yu. Klimenko, Preprint No. 2987, IAÉ (Inst. of Atomic Energy, Moscow, 1978).
41. S. Glasstone, K. J. Laidler, and H. Eyring, *The Theory of Rate Processes* (McGraw-Hill, New York, 1941; Inostrannaya Literatura, Moscow, 1948).
42. P. Anderson, in *Superconductivity*, Ed. by R. D. Parks (Marcel Dekker, New York, 1969), Vol. 2, p. 1343.
43. D. Dew-Hughes, Cryogenics **28**, 674 (1988).
44. Y. Yeshurun, A. P. Malozemov, and A. Shavlov, Rev. Mod. Phys. **68**, 911 (1996).
45. A. V. Mitin, Zh. Éksp. Teor. Fiz. **93**, 590 (1987) [Sov. Phys. JETP **66**, 335 (1987)].
46. A. Gurevich and H. Kupfer, Phys. Rev. B **48**, 6477 (1993).
47. A. Gurevich and E. Brandt, Phys. Rev. Lett. **73**, 178 (1994).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 100, No. 1, 2005, pp. 66–76.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 1, 2005, pp. 74–86.

ORDER, DISORDER, AND PHASE TRANSITIONS IN CONDENSED SYSTEMS

The Magnetic Properties of Eu Nanoclusters Formed in NaCl Crystals during Plastic Deformation and Aggregation

A. A. Baskakov[a], L. R. Dunin-Barkovskiĭ[a,b], R. B. Morgunov[a], Y. Tanimoto[b], and S. Z. Shmurak[a]

[a]*Institute of Solid-State Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*
[b]*Hiroshima University, Higashi-Hiroshima 739-8526, Japan*
e-mail: morgunov@issp.ac.ru

Received June 30, 2004

Abstract—Plastic deformation of aged crystals and diffusion-controlled aggregation of Eu^{2+} ions in freshly quenched NaCl : Eu crystals causes the formation of complexes bound by exchange interaction. The solution of large-sized precipitates by quenching followed by impurity aggregation results in the nucleation of magnetically sensitive clusters comprising few atoms. A constant magnetic field with induction 5 T increases the probability of the transition of magnetically sensitive clusters from the high- to low-spin metastable state and induces changes in their atomic structure. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Studies of magnetoplastic effects in ionic and covalent crystals have led to the development of new physical principles for controlling their mechanical properties with the use of a magnetic field, which changes spin dynamics in separate pairs of defects [1–3]. It was found that some spin-dependent stages of the interaction of paramagnetic defects could proceed at a higher rate than spin-lattice relaxation. They can therefore be sensitive to comparatively low magnetic fields with induction $B \sim 1$ T at near-room temperatures. Part of the magnetoplastic effects were explained by magnetic field-induced transformations of the atomic structure of impurity atom nanoclusters, which played the role of dislocation stoppers [3]. The diffusion-controlled formation of such nanoclusters from separate impurity-vacancy dipoles was intensely studied in [4–10], but possible influence of electron-spin processes on the aggregation of impurities was not taken into account. Studies of magnetic field effects on cluster formation in a crystal lattice can provide new information about elementary plastic deformation processes and the possibility of controlling them with the use of magnetic fields. In addition, they can also contribute to distinguishing the contribution of electron-spin processes to impurity aggregation.

It was shown in [11–14] that the atomic structure of clusters formed during plastic deformation or the aggregation of Eu^{2+} ions in the crystal lattice of NaCl could be changed by applying a magnetic field and that these changes could be observed in photoluminescence spectra. Understanding the micromechanisms of this process requires data on the spin state of clusters before and after their transformation in a magnetic field. The general principles of the physics of nonequilibrium spin-dependent phenomena [15–20] present a single possibility for explaining the influence of magnetic fields on the atomic structure of clusters: changes in their spin state at the instants of their excitation by thermal fluctuations. The initial and at least one of the possible final states of such clusters should correspond to stable spin-correlated pairs, triples, or more complex systems with exchange interaction, which maintains spin correlation in clusters at high temperatures. A magnetic field is capable of changing the mutual orientation in clusters only at the instants of the transition from one stable state to another, when the energy difference between them is comparable to the Zeeman splitting of spin levels and is on the order of $\mu_B gB$, where B is the constant magnetic field induction, μ_B is the Bohr magneton, and g is the g factor of the paramagnetic particle. The magnetic properties of clusters sensitive to magnetic fields, their initial and final spin states, and the contribution to the total magnetic susceptibility of the paramagnetic impurity subsystem have not been determined thus far for the systems characterized by magnetoplastic effects or the influence of a weak magnetic field on their other physical properties.

Our goal was (1) to study the magnetic properties of Eu clusters formed as a result of plastic deformation of NaCl crystals and diffusion-controlled aggregation of impurity-vacancy Eu dipoles and (2) detect magnetic field effects on the effective spin of magnetically sensitive nonequilibrium Eu clusters and determine the direction of the corresponding transition.

1063-7761/05/10001-0066$26.00

2. EXPERIMENTAL

We used NaCl : Eu crystals of about 0.1 g containing 600–1000 ppm europium. The crystals were grown by the Czochralski method in a controlled dry argon atmosphere. Doping was performed by adding a certain amount of $EuCl_2$ preliminarily reduced from $EuCl_3 \cdot 6H_2O$ to NaCl melts.

The real part of magnetic susceptibility was measured using an MPMS-7 SQUID magnetometer in a constant magnetic field (0.3 T) over the temperature range 2–300 K. The sensitivity of the magnetometer with respect to the magnetic moment under constant magnetic field conditions in the reciprocating sample option (RSO) mode was 10^{-8} emu, which was sufficient for the determination of magnetic moments M with an accuracy on the order of 10^{-4}. Our main goal was to study the paramagnetic (or superparamagnetic) component of the magnetic moment of the samples related to the contribution of Eu^{2+} ions. For noninteracting Eu^{2+} ions, the $M(T)$ dependence obeys the Curie law; that is, $M(T)T$ = const at high temperatures and begins to decrease according to the Brillouin function at low temperatures ($kT \sim \mu_B gB$), because the distribution of paramagnetic particles over spin states then changes substantially. Another reason for a decrease in $M(T)T$ as temperature lowers can be negative exchange interaction within europium clusters. Positive exchange interaction can increase $M(T)T$ compared with a system of noninteracting spins. The influence of weak exchange interaction can be detected at fairly low temperatures $kT \sim \Theta$, where Θ is the exchange interaction energy.

As the filling of spin states and exchange interaction in clusters are the main factors that control their magnetic sensitivity, the low-temperature region of the $M(T)$ dependence is of the greatest interest. Against the background of the Curie law, the effect of admixtures related to exchange interaction is usually weak. For this reason, the contribution of exchange interactions and the spin of particles were determined by applying the usual approach; that is, we constructed the dependence of the product $M(T)T$ on temperature and other experimental parameters. Whereas the crystal was subjected to mechanical and magnetic actions at room temperature, the magnetic state of the impurity was determined using the most informative low-temperature portion of the $M(T)T$ dependence.

In all our crystals, the $M(T)T$ dependence in the region of high temperatures was linear and deviated from the straight line only at $T < 20$–30 K (Fig. 1a). This is explained by a substantial contribution of the crystal lattice, whose diamagnetic moment M_{lat} does not depend on temperature, to the total magnetic moment of the sample $M = M_{lat} + M_{Eu}$. We therefore subtracted the diamagnetic component from the experimental temperature dependences $M(T)T$ by the standard method [21]. The diamagnetic component was determined from the slope of the $M(T)T$ dependence at near-room temperatures, where $T \gg \Theta$ and the contribution of the $M_{Eu}T$ paramagnetic component did not depend on temperature. The slope of the experimental $M(T)T$ dependence was determined via approximating its high-temperature portion by a linear function. The $M_{Eu}(T)T$ dependences obtained after subtracting the diamagnetic component (Fig. 1b) characterize the contribution of only the paramagnetic (or superparamagnetic) M_{Eu} component of the magnetic moment of the crystal. We also determined the M_{lat} value by weighing the sample and calculating the magnetic moment from the reference data on the diamagnetic susceptibility of NaCl. This method gave the same M_{lat} value as that obtained from the temperature dependence $M(T)T$.

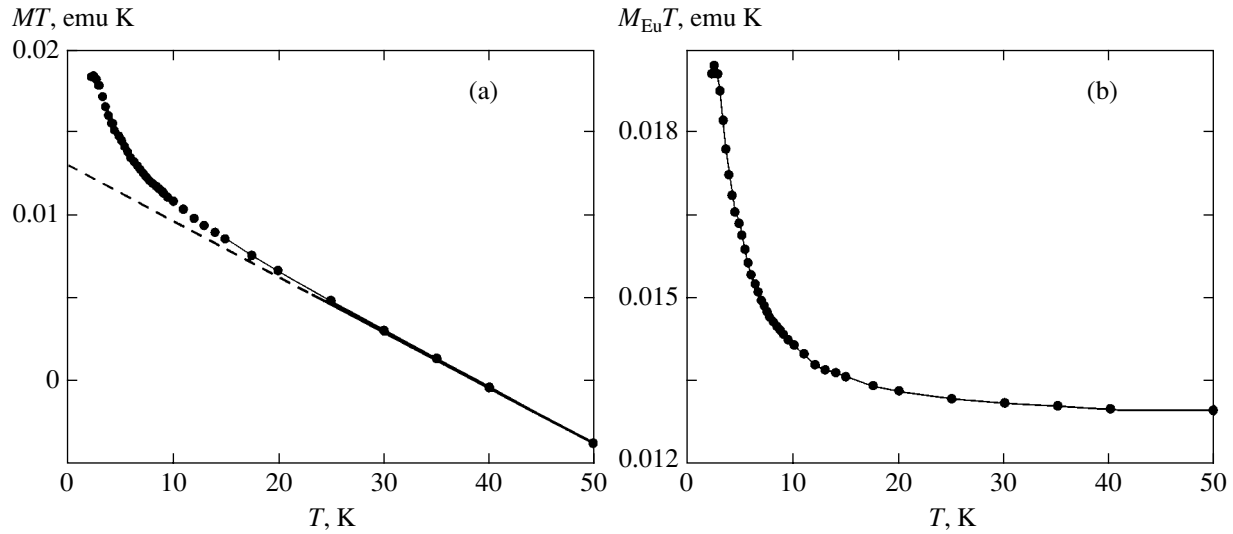


Fig. 1. Temperature dependences of magnetic moment M multiplied by temperature T for a NaCl : Eu crystal (a) before subtracting the diamagnetic component and (b) after its subtraction.

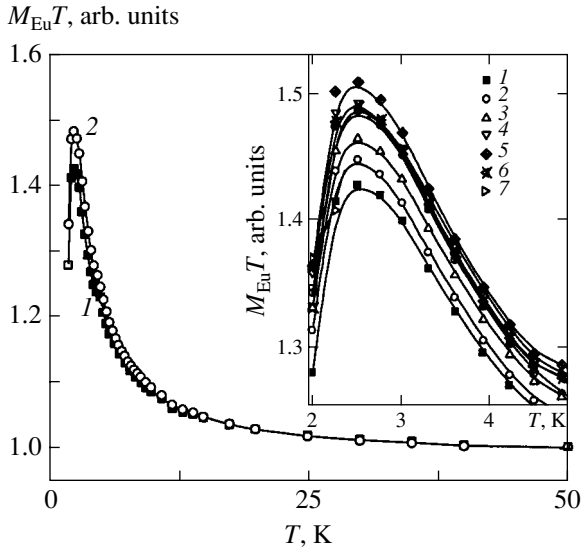


Fig. 2. Temperature dependence $M_{Eu}(T)$ in crystals with deformations (*1*) $\varepsilon = 0$ and (*2*) $\varepsilon = 10\%$. The $M_{Eu}(T)$ value at room temperature is taken to be unity. Solid lines are the calculated $M_{Eu}(T)T$ dependences obtained including the contributions of isolated impurity-vacancy dipoles and exchange-coupled clusters. The low-temperature portions of the $M_{Eu}(T)T$ dependences at various deformations ε = (*1*) 0, (*2*) 2, (*3*) 4.8, (*4*) 5.6, (*5*) 7, (*6*) 8, and (7) 10% are shown in the inset.

Special precautions were taken during the preparation of the samples. After chopping them out or thermal treatment, a surface layer approximately 100 μm thick was removed by polishing with purified deionized water (ultrapure water 43001-1B) to prevent the magnetic properties of crystals from being influenced by accidentally present foreign ferromagnetic particles. The crystals were handled exclusively with ceramic pincers and deformed in a deformation machine with quartz rods. This prevented the samples from being contaminated with foreign magnetic particles that might be pressed into their surface.

The crystals were subjected to macroplastic compression deformation on a "soft" deforming machine, in which mechanical stress linearly grew with time, $\sigma \propto t$, and the relative deformation of the sample ε was measured. Industrial "hard" deformation machines of the Instron type are often used in experiments. In such machines, a linear time sweep is set for the deformation $\varepsilon \propto t$ rather than mechanical stresses. The deformation is proportional to the mean rate of an increase in the area "swept" by dislocations. The rate of deformation was not constant in our experiments; we can only specify its mean value, which was on the order of 10^{-5} s^{-1}. The photoluminescence of the crystals was studied at room temperature using a Jasco-777 standard spectrophotometer. The excitation light had a wavelength of $\lambda = 370$ nm, and the spectra were recorded over the wavelength range $\lambda = 380$–735 nm. The EPR spectra were obtained on a Bruker ESP-300 spectrometer at $T = 3.3$ K (microwave field frequency 9.453 GHz). The constant magnetic field of the spectrometer was directed along the (001) axis. In several experiments, we used a JMTD-LH15T40 superconducting laboratory magnet, which enabled the samples to be subjected to the action of a constant magnetic field up to 15 T at room temperature.

3. RESULTS

3.1. The Influence of Plastic Deformation on the Magnetic and Optical Properties of Eu Clusters

The influence of plastic deformation on the magnetic properties of clusters was studied for crystals held for five years at room temperature and not subjected to thermal treatment. The major fraction of impurity ions were therefore in the state of large-sized precipitates. The crystals contained 1000 ppm europium. After the first measurement of the temperature dependence of the MT product, the NaCl : Eu samples were placed into a deforming machine and loaded with a mechanical load that linearly increased with time. The dependence of the relative crystal deformation ε on mechanical stress σ was recorded. After $\varepsilon = 1$–2% was reached, the crystals were unloaded and the residual plastic deformation ε_{res} was determined using a micrometer. Next, the samples were placed into the SQUID magnetometer, and the dependence $M_{Eu}(T)T$ was recorded. The crystals were then again deformed and the $M_{Eu}(T)T$ dependence recorded. We performed several deformations and measurements of the temperature dependence of the magnetic moment for every sample. We found that plastic deformation changed the shape of the $M_{Eu}(T)T$ dependence (Fig. 2). At deformations exceeding 5–6%, which corresponded to the hardening stage, further deformation of the samples did not cause changes in the $M_{Eu}(T)T$ dependences (Fig. 2).

It was reported in [22] that the magnetic moment of chemically "pure" NaCl crystals changed after their deformation and that there remained residual magnetization whose temperature dependence was characteristic of ferromagnets. To separate the contributions of crystal lattice deformation and changes in the state of the paramagnetic impurity subsystem, we performed experiments with NaCl : Cu crystals. Deforming crystals with a copper concentration of about 1 ppm caused the crystal magnetic moment to change by about 10^{-8} emu, which was close to the instrumental error of the magnetometer that we used and three to four orders of magnitude smaller than changes that we observed in NaCl : Eu crystals after their deformation. It follows that, in our experiments, all changes in the magnetic properties of the NaCl : Eu crystals resulting from plastic deformation were caused by processes that occurred in the subsystem of the paramagnetic europium impurity.

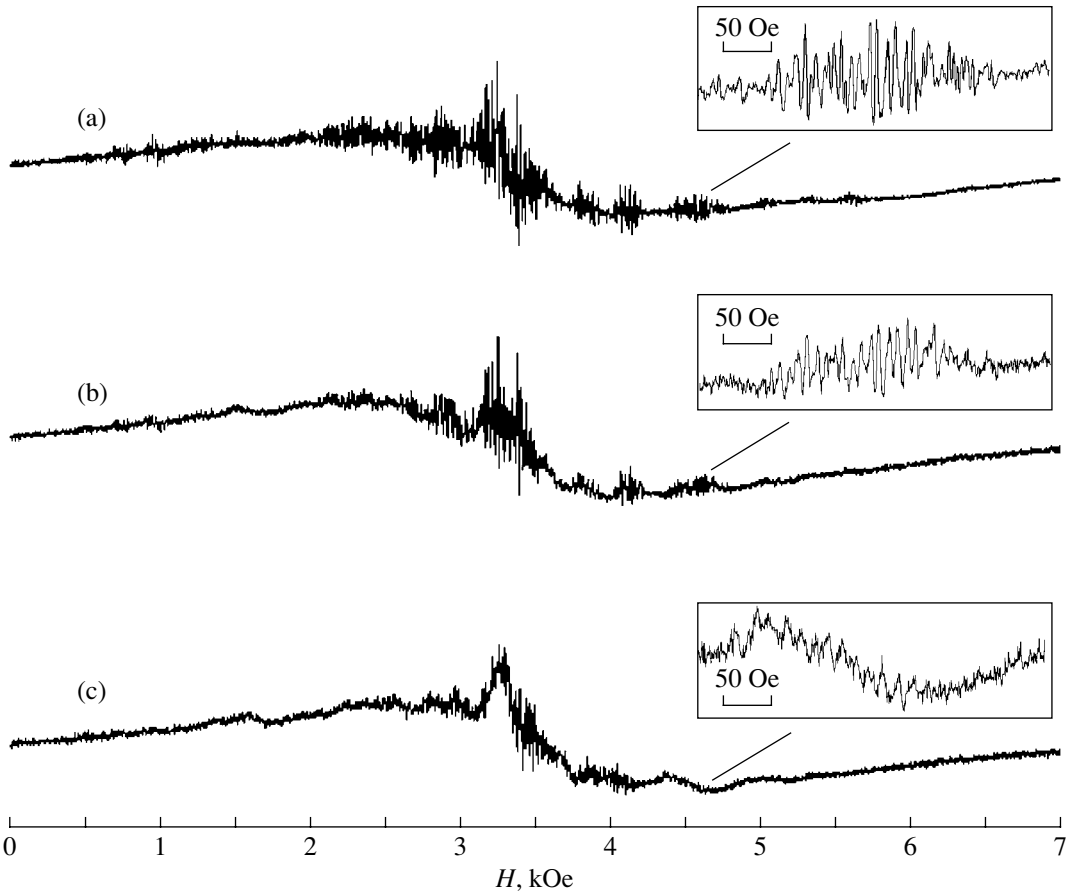


Fig. 3. EPR spectra of an aged NaCl : Eu crystal (a) before deformation, (b) after deformation to ε = 6%, and (c) after deformation to ε = 12%. Spectrum fragments containing narrow lines of impurity-vacancy dipoles are shown in insets.

The EPR spectra of the samples recorded at 3.3 K before plastic deformation and after deforming them by 6 and 12% are shown in Fig. 3. The fragments of these spectra given in the insets contain series of narrow lines about 5 Oe wide. The sequence of the spectra presented in Fig. 3 shows that the fraction of narrow lines decreases and several comparatively broad bands appear as the degree of plastic deformation grows.

Apart from changes in the magnetic properties, deforming the crystals caused the redistribution of intensities in various regions of the photoluminescence spectra of the NaCl : Eu crystals and the appearance of a new band at λ = 536 nm observed by us in [23] (Fig. 4). As with deformation in a hard machine [23], the luminescence spectrum remained almost unchanged up to the elastic limit. After the yield point was reached, when the density of dislocations increased substantially, we observed a decrease in the intensity of luminescence at λ = 427 nm and its increase at λ = 536 nm (Fig. 4). At the end of the stage of easy glide of dislocations (at ε = 5–6%), these changes slowed down, and, at the hardening stage, the intensities of luminescence in these spectral regions almost ceased to depend

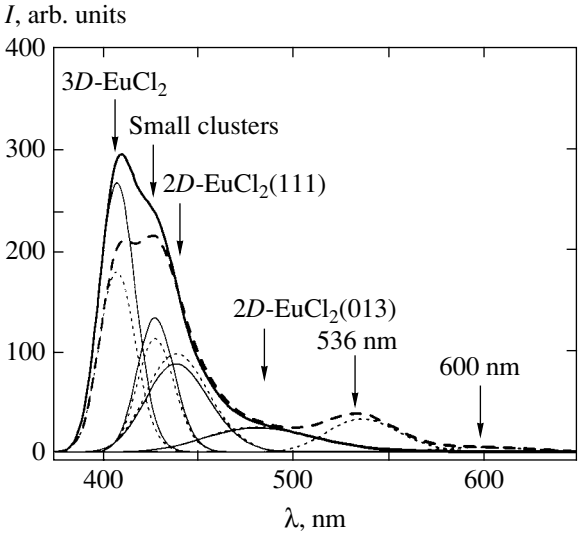


Fig. 4. Photoluminescence spectra of an aged NaCl : Eu crystal before deformation (thick solid line) and after plastic deformation to ε = 10% (dashed line). The thin solid and dotted lines are the corresponding spectra resolved into Gauss profiles before and after deformation under the conditions of mechanical stress increasing linearly. Arrows indicate cluster types corresponding to each spectral band according to [4].

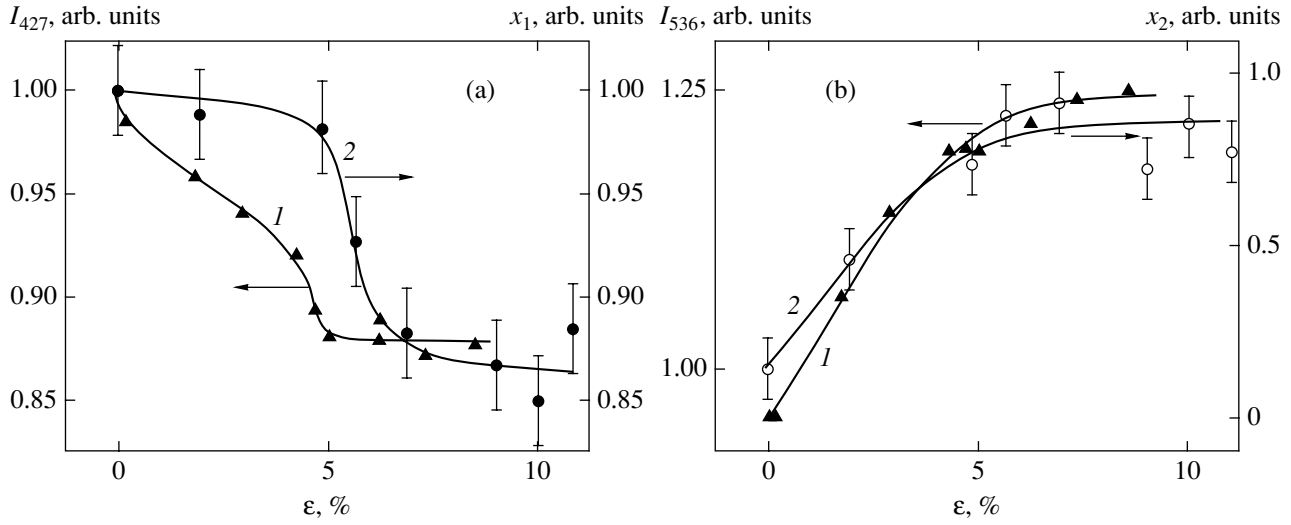


Fig. 5. (a) Deformation ε dependences of (*1*) luminescence intensity I_{427} of impurity-vacancy dipoles and (*2*) number x_1 of impurity-vacancy dipoles determined by analyzing SQUID measurements; the I_{427} and x_1 values before deformation were taken to be unity. (b) Deformation ε dependences of (*1*) luminescence intensity I_{535} of clusters (bands centered at 536 nm) and (*2*) number x_2 of exchange-coupled complexes determined by analyzing SQUID measurements; the I_{535} and x_2 values before deformation were taken to be unity.

on further plastic deformation (Fig. 5). It follows that the magnetic properties of the crystals and changes in the luminescence spectrum initiated by deformation in a soft machine are sensitive to changes of the crystal deformation stages and the presence and regime of dislocation movement.

3.2. Changes in the Magnetic Properties of Crystals during the Aggregation of Impurity-Vacancy Dipoles into Clusters

Changes in the magnetic moment M_{Eu} caused by thermal treatment of the NaCl : Eu (600 ppm) crystals and impurity aggregation were studied after holding the samples at 770 K for 2 h and throwing them onto a copper plate to increase the rate of cooling. This procedure (quenching) caused strong changes in the $M_{Eu}(T)T$ curve. The dependence characteristic of superparamagnets acquired the shape typical of paramagnets after quenching (Fig. 6). These changes were caused by the solution of large-sized clusters bound by exchange interaction and their transformation into isolated impurity-vacancy dipoles. The low-temperature tail of the $M_{Eu}(T)T$ dependence changed after quenching as time elapsed (Fig. 7). The $M_{Eu}(T)T$ dependence showed a tendency to return to its initial form (to the form it had before quenching) at room temperature because of the gradual diffusion-controlled aggregation of impurity-vacancy dipoles into clusters.

We also observed changes in the luminescence spectra of the crystals caused by the aggregation of impurity-vacancy dipoles (Fig. 8). These were an increase in the intensity of long-wave spectrum components and a decrease in the integral intensity of luminescence. For four pairs of samples whose luminescence spectra were absolutely identical immediately after quenching, we observed spectral differences long after, depending on whether or not the crystal was held in a constant field of 15 T. The intensity of long-wave luminescence of the samples that remained in a magnetic field for the first 50–100 h after quenching was higher compared with control samples. This observation is in agreement with the results reported by us in [11–14]; it means that magnetically sensitive clusters whose transformation

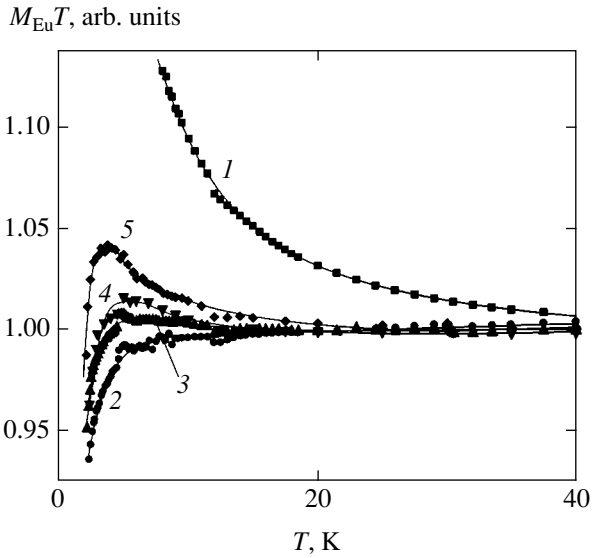


Fig. 6. Temperature dependences of $M_{Eu}T$ for an aged crystal (*1*) before quenching and (*2*) 5 h, (*3*) 30 days, (*4*) 60 days, and (*5*) 120 days after quenching.

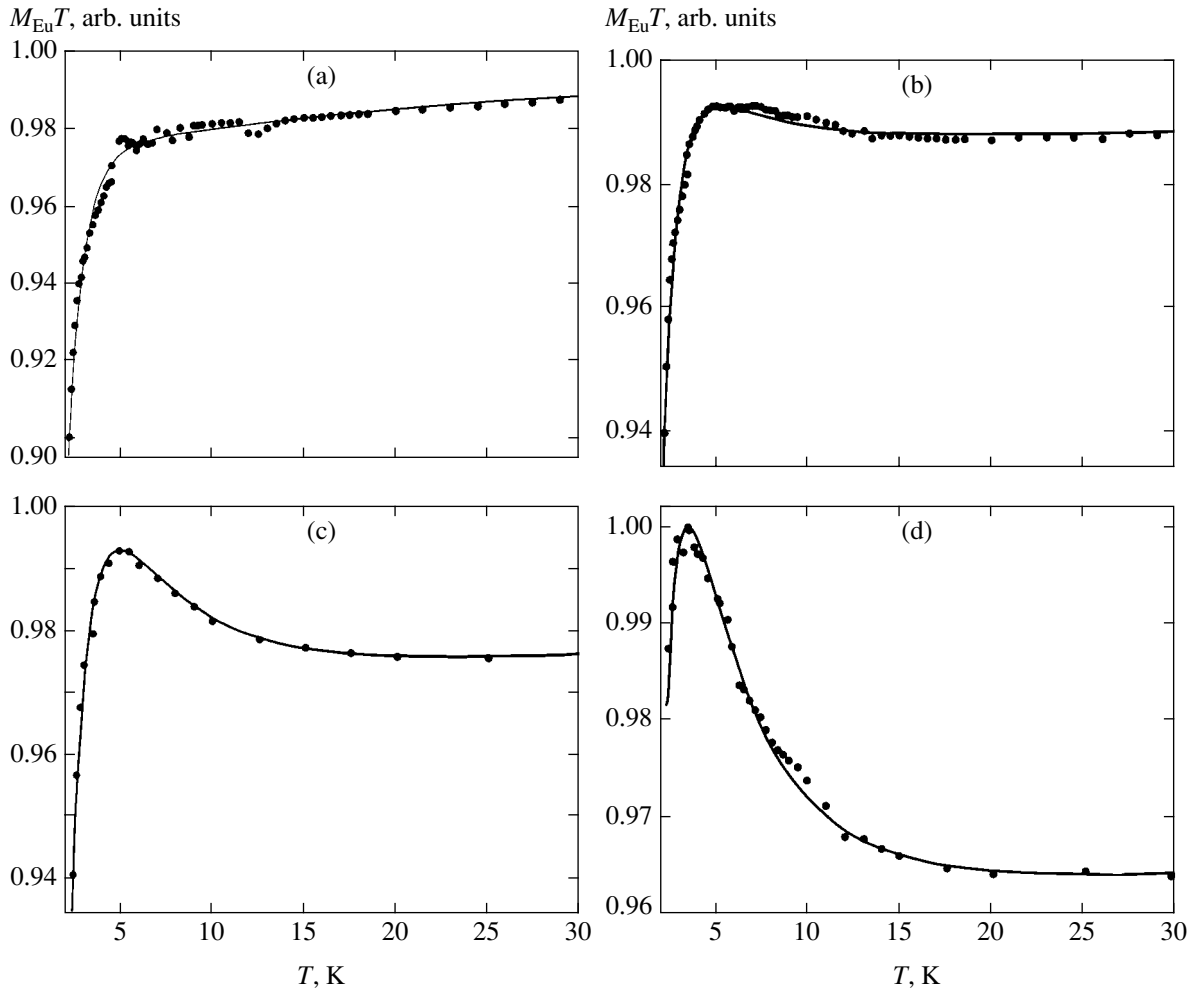


Fig. 7. Low-temperature regions of $M_{Eu}(T)T$ dependences for quenched crystals (a) 5 h, (b) 30 days, (c) 60 days, and (d) 120 days after quenching. Solid lines are the calculated $M_{Eu}(T)T$ dependences obtained taking into account the contributions of isolated impurity-vacancy dipoles and dimers in various spin states.

causes "switching" between aggregation paths appear during aggregation. The next series of experiments were therefore performed to study these magnetically sensitive nonequilibrium states.

3.3. The Influence of a Constant Magnetic Field on the Effective Spin of Magnetically Sensitive Nonequilibrium Europium Clusters

It was shown in [11–14] that magnetically sensitive clusters comprising few atoms appeared approximately 50 h after quenching if aggregation occurred at room temperature. In other words, a field with induction of about 6 T influences the microhardness of crystals and photoluminescence if intermediate clusters of impurity-vacancy dipoles are formed in crystals; the concentration of such dipoles is maximum 50 h after quenching. A magnetic field with induction less than 1 T almost does not change the properties of crystals. It was found [11, 12] that a decrease in the temperature to 77 K suppressed the magnetoplastic effect and the field effect on photoluminescence. Taking all these observations into consideration, we used a weak measuring field with induction of 0.3 T and a strong magnetic field capable of transforming clusters of point defects. The sequence of operations is schematically shown in Fig. 9. After quenching, the NaCl : Eu (600 ppm) crystals were held at room temperature for two days, and the $M_{Eu}(T)T$ dependence was then measured in a weak magnetic field from 2 to 200 K. Every crystal was heated to 300 K and held for 40 min in the absence of external actions. This was necessary for determining those changes in M that might be caused by impurity aggregation. It was then cooled to 2 K, and the $M_{Eu}(T)T$ dependence was again measured. After heating to 300 K for the next time, a strong magnetic field with an induction of 5 T was switched on in the SQUID magnetometer, and the sample was held in this field for 40 min. Magnetic field-induced changes in the crystal

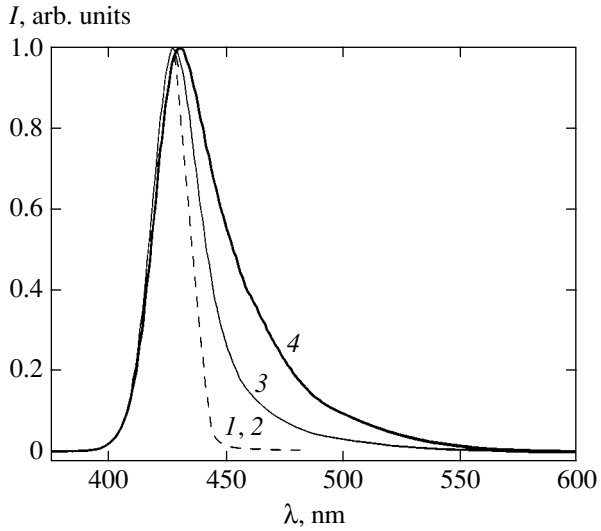


Fig. 8. Photoluminescence spectra of two quenched NaCl : Eu crystals: (*1*) and (*2*) coinciding spectra immediately after quenching, (*3*) spectrum of a control sample held 6 months after quenching at room temperature in the absence of a magnetic field, and (*4*) spectrum of the sample held in a field of 15 T for the first 100 h after quenching and stored 6 months after quenching at room temperature.

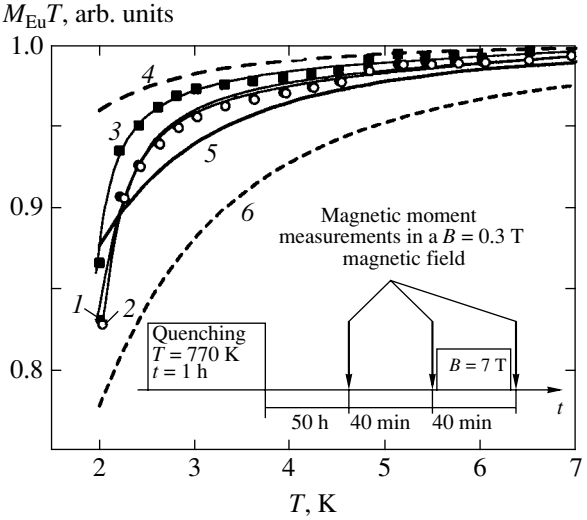


Fig. 9. Experimental $M_{Eu}(T)T$ dependences: (*1*) the first series of measurements prior to exposure to a 5 T field, (*2*) the second series of measurements after heating before exposure to a 5 T field, and (*3*) the third series of measurements after exposure to a 5 T field, and calculated $M_{Eu}(T)T$ dependences obtained using the Brillouin function that describes (*4*) isolated dipoles with spin $S = 7/2$, (*5*) dimers with the highest spin possible $S = 7$, and (*6*) trimers with the highest spin possible $S = 21/2$. The sequence of operations is shown in the inset. Arrows correspond to $M_{Eu}(T)T$ dependence measurements in a 0.3 T field.

were studied by again recording the $M_{Eu}(T)T$ dependence in a weak field. Importantly, the sample remained in the magnetometer during all these manipulations, and its accidental displacements, inevitable if the sample were taken out, could not influence the experimental results.

These results are shown in Fig. 9. The low-temperature tail of the $M_{Eu}(T)T$ dependence remained almost unchanged when the samples were heated. The exposure of the crystals to a 5 T field caused changes in the form of the $M_{Eu}(T)T$ dependence at low temperatures. These changes were reproduced for all six samples used in this series of experiments. We checked the possibility of the influence of extraneous factors on the purity of our experiment. The rated residual magnetic field in the magnetometer after switching the 5 T field off was about 0.1 Oe, that is, on the order of 10^{-6} of the main signal. The presence of such a weak additional field cannot explain changes in the magnetic moment of the crystal observed in our experiments. We nevertheless performed control measurements with NaCl : Cu(1–10 ppm) crystals that were not subjected to thermal treatment. Because of the absence of excited states of the subsystem of paramagnetic impurities, it was impossible in principle to observe changes induced by a magnetic field at room temperature. In other words, these samples were used as a field-insensitive reference. The $M(T)T$ dependence did not change after the exposure of the NaCl : Cu crystals to field action (Fig. 10). It follows that the residual field in the magnetometer, traces of ferromagnetic particles on the surface of the crystals, and other artifacts could not cause changes in the magnetic properties of the NaCl : Eu crystals observed in our experiments.

4. DISCUSSION

First, let us consider changes in the magnetic properties of the crystals caused by their deformation. The Eu^{2+} ion has spin $S = 7/2$. In the absence of exchange interactions, its EPR spectrum comprises fourteen packets of lines about 5 Oe wide split by hyperfine interactions [23]. Clearly, these narrow lines characterize the fraction of isolated impurity-vacancy dipoles in our experiments. The broad lines are the EPR spectrum of clusters broadened by dipole or exchange interaction (see Fig. 3). Changes in the EPR spectrum while the crystal is deformed (a decrease in the fraction of narrow lines and the appearance of broad lines) are evidence of a decrease in the fraction of isolated impurity-vacancy dipoles and the appearance of new clusters (see Fig. 3). As the aged crystals contained clusters of various types, an analysis of the SQUID magnetometry data on exchange interactions within these clusters could only be used to obtain averaged estimates taking into account the corresponding correction to the Brillouin functions [24]. The $M(T)T$ dependence was therefore approximated by the additive contributions of x_1 isolated impurity-vacancy dipoles M_{dip} and x_2 exchange-coupled clusters M_{cl} to the magnetic moment. The temperature dependence $M_{dip}(T)$ was described by the Brillouin function $B_S(\zeta)$, which took into account the Bolt-

zmann spin-level distribution of electrons as a function of temperature,

$$M_{\text{dip}}(T) = g\mu_B S x_1 B_S(\zeta), \quad (1)$$

where

$$B_S(\zeta) = \frac{2S+1}{2S}\coth\frac{(2S+1)\zeta}{2S} - \frac{1}{2}S\coth\frac{\zeta}{2S},$$

$$\zeta = \frac{Sg\mu_B B}{kT},$$

μ_B is the Bohr magneton, and $g \approx 2$ is the g factor.

The M_{cl} contribution of exchange-coupled clusters was described by a similar function where the argument of the Brillouin function was changed taking into account the correction T_0 for averaging the exchange interaction value $J \sim kT_0$ over various clusters,

$$\zeta = Sg\mu_B B/k(T + T_0).$$

To summarize, the experimental $M(T)T$ dependences were approximated by the function $M_{\text{Eu}}(T) = M_{\text{dip}}(T) + M_{\text{cl}}(T)$ whose adjustment parameters were x_1, x_2, and T_0. Figure 2 shows that the calculated curve closely agrees with the experimental dependences at $T_0 = -2$ K. As estimating dipole–dipole interactions yields $T_0 \sim 0.1$ K, this result characterizes mean exchange antiferromagnetic interactions in clusters. The above approximation of the experimental data enabled us to find the dependences of the number of isolated impurity-vacancy dipoles x_1 (Fig. 5a) and exchange-coupled clusters x_2 (Fig. 5b) on the deformation ε of crystals. In conformity with the conclusions obtained in analyzing the photoluminescence and EPR spectra, an analysis of the SQUID magnetometry data shows that the number of isolated dipoles decreases and that of exchange-coupled clusters increases as the crystals are deformed. At the stage of hardening, the $x_1(\varepsilon)$ and $x_2(\varepsilon)$ dependences reach saturation. This conclusion, substantiated by three independent methods, contradicts the result obtained when the same crystals were deformed using a hard machine [25]. It was shown in [25] that clusters were destroyed and transformed into isolated impurity-vacancy dipoles or their complexes containing few atoms under the conditions of plastic deformation at a constant rate in a hard machine. This means that the regime of deformation substantially influences the character of interactions between dislocations and point defects. The aggregation of clusters during plastic deformation observed in this work can be explained by an increased mobility of impurities along the dislocation nucleus and an ensuing increase in the probability of the approach of dipoles to each other in this region and also by the entrainment of impurities with moving dislocations. Drawing conclusions in favor of one of these possibilities requires performing additional experiments.

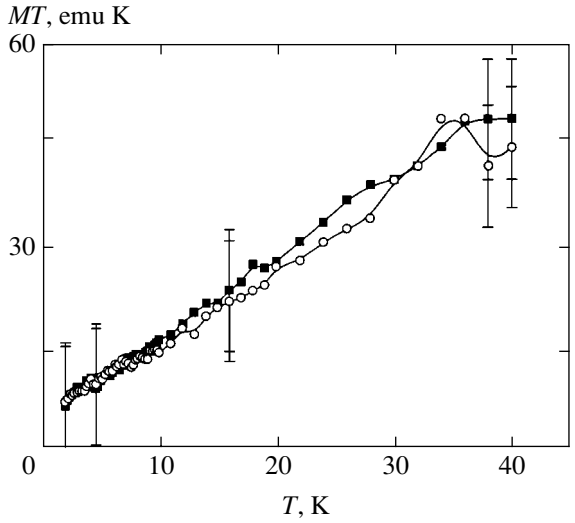


Fig. 10. Dependences $M(T)T$ for aged NaCl : Cu(1 ppm) crystals obtained (*1*) before and (*2*) after exposure to a 5 T field.

Turning to the influence of thermal treatment on the magnetic properties of europium impurities, note that the EPR spectrum observed in our experiments after quenching corresponded to the presence in crystals of only isolated impurity-vacancy complexes or their weakly bound clusters, dimers. Various stages of impurity aggregation in ionic crystals were studied theoretically in [26]. It was shown that the primary products of diffusion-controlled aggregation, that is, dimers (pairs of dipoles), could be of four types in the NaCl lattice (Fig. 11), and that all configurations formed equiprobably at the initial stages because they appeared in accidental encounters of dipoles wandering in the lattice under the action of thermal fluctuations. This stage is the "bottleneck" of further aggregation, because only two of the four dimer configurations admit the addition of dipoles and cluster growth. It follows that unfavorable configurations should first be transformed into favorable dimers by thermal fluctuations for trimers and more complex formations to appear. This conclusion is substantiated by the kinetic studies of changes in the EPR spectra after quenching [13].

In this work, we studied comparatively short initial stages of aggregation at which large-sized clusters could not be formed. At a 600 ppm concentration of impurity, the mean distance between dipoles is $d \sim 100$ Å. Therefore, the diffusion coefficient being $D \sim 10^{-22}$ m^2/s, the drawing of two dipoles together takes time $t \sim d^2/D \sim 10^7$–10^8 s, that is, about 100 days. Our

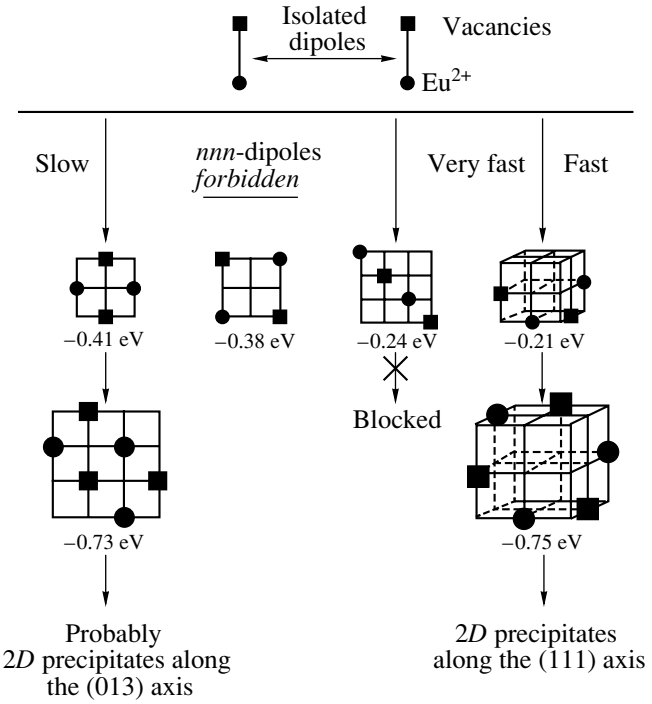


Fig. 11. Schematic representation of aggregation paths from dipoles to clusters according to [26].

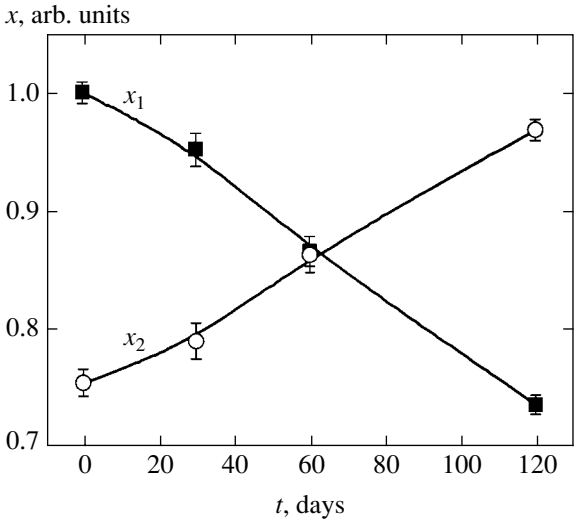


Fig. 12. Dependences of the numbers of dipoles (x_1) and dimers (x_2) on time elapsed after quenching NaCl : Eu crystals.

measurements were performed approximately during this time interval. For this reason, the dependence $M_{Eu}(T)T$ was modeled taking into account the contributions of x_1 isolated impurity-vacancy dipoles and x_2 dimers of various configurations, $M_{Eu} = M_{dip} + M_{dim}$. The contribution of dipoles M_{dip} was described by (1), and the contribution of dimers M_{dim} had to be written taking into account all possible spin states. The exact equation for the magnetic moment of dimers was taken from [27],

$$M_{dim} = \frac{1}{2}x_2 g\mu_B \times \frac{\sum_{S=0}^{S_{max}} \exp\left(\frac{JS(S+1)}{kT}\right)\sinh\left(\frac{2S+1}{2S}\zeta\right)SB_S(\zeta)}{\sum_{S=0}^{S_{max}} \exp\left(\frac{JS(S+1)}{kT}\right)\sinh\left(\frac{2S+1}{2S}\zeta\right)}, \quad (2)$$

where $\zeta = Sg\mu_B B/k(T + T_0)$.

The approximation of the experimental data at various times after quenching by the function

$$M_{Eu}(T)T = [M_{dip}(T) + M_{dim}(T)]T$$

is shown above in Fig. 7. We see that, during the whole time interval of measurements, the experimental data are well described by the model that takes into account the contributions of dipoles and dimers. This substantiates the correctness of the selected model although does not rule out the formation of more complex clusters, which, it appears, do not make a substantial contribution to the magnetic moment of the crystal. Fitting the $M_{Eu}(T)T$ dependences with the use of the above function enabled us to determine the amounts of dipoles (x_1) and dimers (x_2) as functions of time elapsed after quenching. Figure 12 shows that aggregation decreases the number of dipoles and increases that of dimers, as is expected for the initial cluster formation stages.

The influence of a magnetic field on the mean spin of the clusters formed during aggregation can be analyzed by comparing the experimental $M_{Eu}(T)T$ curves with the theoretical predictions obtained with the use of the Brillouin function corresponding to the states with maximum spin S_{max} of dipoles and clusters of various types, including dimers, trimers, etc. It follows from Fig. 9 that the experimental curves lie between the Brillouin functions written for dipoles and for the high-spin state of dimers with $S = 7$. Using the mean spin as an approximation parameter in the Brillouin function yields an exact description of the $M_{Eu}(T)T$ dependence before and after applying a strong constant magnetic field (Fig. 9). This approach allowed us to find that the mean spin of clusters before and after applying the field was 6.2 ± 0.1 and 4.8 ± 0.1, respectively. It follows that the exposure of the crystals to field action decreases the mean spin, which can be explained by the transition of part of clusters from the high- to low-spin state. As SQUID magnetometry only allows the total contribution of all types of clusters to magnetization to be measured, the absolute spin value of magnetically sensitive clusters cannot be determined by this method. It, however, follows from very general considerations that spins in a cluster should be coupled by strong exchange

interaction for thermal fluctuations to be incapable of rapidly destroying such complexes and spin correlation to persist for a long time. The closeness of the mean spin before applying a magnetic field, $S = 6.2$, to its maximum value in dimers, $S_{max} = 7$, therefore leads us to suggest that magnetically sensitive clusters are dimers (pairs of dipoles) in which spins are coupled by exchange interaction and oriented parallel to each other.

Note in conclusion that the aggregation of particles in crystal lattices has currently been studied in detail only at late stages of its development, when combining a large number of particles results in the formation of nanocrystals discernible in electron microscopic images and X-ray diffraction patterns. At the same time, the initial stages of formation of nuclei of cluster growth are crucial to the process as a whole. These clusters only comprising several impurity atoms can have different atomic configurations and be characterized by different electronic processes involved in their origination. It could be suggested that the physical mechanisms of formation of nuclei are fairly general for a wide range of crystals. The information obtained in this work can be used to solve such problems of importance for applications as the doping of semiconductors and controlling the state of clusters in them, crystal growth, and the self-organization of impurity atoms in crystal lattices of metals and alloys. Aggregation is usually considered from the point of view of jumps of atoms in crystal lattices under the action of thermal fluctuations. The corresponding theoretical concepts are largely based on calculations of lattice dilatancy and its influence on the diffusion coefficient. Other important characteristics of aggregation, namely, the occurrence of electronic and spin reactions between paramagnetic particles with spins, are then outside the field of vision. In this sense, aggregation is akin to processes of concern to spin chemistry, in which similar reactions were only studied for liquid-phase systems [15, 16]. We can expect considerable changes in the physics of spin-dependent reactions in solids compared with processes in liquids. Dipole–dipole and spin–orbit interactions, intracrystalline field anisotropy, and different molecular dynamics of paramagnetic particles change the effectiveness of solid-state spin-dependent processes. These problems require additional inquiries.

5. CONCLUSIONS

(1) The plastic deformation of crystals under the conditions of mechanical stress linearly increasing with time was found to decrease the concentration of isolated impurity-vacancy dipoles and cause the formation of clusters with exchange interaction of the ferromagnetic type. This can be caused by the arising of conditions favoring the approach of dipoles to each other in regions of strong crystal lattice distortions close to dislocation nuclei and the entrainment of dipoles by moving dislocations.

(2) Aging crystals at room temperature for 10^3 h caused aggregation of isolated impurity-vacancy dipoles to exchange-coupled clusters (dimers) containing two Eu^{2+} ions. The temperature dependences of the magnetic moment of clusters can be correctly described without including the contribution of larger aggregates.

(3) Among the clusters formed approximately 50 h after quenching, there were magnetically sensitive configurations that changed their optical and magnetic properties at room temperature under the action of a magnetic field with induction 5 T. This field was found to initiate the transition of part of clusters from the high- to low-spin state.

ACKNOWLEDGMENTS

This work was financially supported by the Presidential Program for Support of Studies of Young Doctors in Sciences (project no. 02-15-99302), the Russian Foundation for Basic Research (project no. 04-02-17576), and the State Program of the Russian Academy of Sciences “Spintronics.”

REFERENCES

1. V. I. Al’shits, E. V. Darinskaya, M. V. Koldaeva, and E. A. Petrzhik, Kristallografiya **48**, 826 (2003) [Crystallogr. Rep. **48**, 768 (2003)].
2. Yu. I. Golovin, Fiz. Tverd. Tela (St. Petersburg) **46**, 769 (2004) [Phys. Solid State **46**, 789 (2004)].
3. R. B. Morgunov, Usp. Fiz. Nauk **174**, 131 (2004) [Phys. Usp. **47**, 125 (2004)].
4. J. Rubio, J. Phys. Chem. Solids **52**, 101 (1991).
5. A. E. Cordero-Borboa, O. Cano-Corona, A. Clevel-Hernandez, and E. Orozco, J. Phys. C **19**, 7113 (1986).
6. J. A. Munoz, E. Rodriguez, J. O. Tocho, and F. Cusso, J. Lumin. **72–74**, 233 (1997).
7. N. M. Bannon and J. Corish, Philos. Mag. A **51**, 797 (1985).
8. F. J. Lopez, H. Murrieta, S. A. Hernandez, and J. Rubio, Phys. Rev. B **22**, 6428 (1980).
9. J. E. Muñoz-Santiuste and J. Garccí-Solé, Phys. Rev. B **38**, 10874 (1988).
10. P. W. M. Jacobs, J. Phys. Chem. Solids **51**, 35 (1990).
11. R. B. Morgunov, A. A. Baskakov, I. N. Trofimova, and D. V. Yakunin, Fiz. Tverd. Tela (St. Petersburg) **45**, 257 (2003) [Phys. Solid State **45**, 270 (2003)].
12. R. B. Morgunov, S. Z. Shmurak, A. A. Baskakov, *et al.*, Pis’ma Zh. Éksp. Teor. Fiz. **76**, 366 (2002) [JETP Lett. **76**, 307 (2002)].
13. R. B. Morgunov and A. A. Baskakov, Fiz. Tverd. Tela (St. Petersburg) **45**, 91 (2003) [Phys. Solid State **45**, 94 (2003)].
14. R. B. Morgunov and A. A. Baskakov, Fiz. Tverd. Tela (St. Petersburg) **43**, 1632 (2001) [Phys. Solid State **43**, 1700 (2001)].

15. B. Ya. Zel'dovich, A. L. Buchachenko, and E. L. Frankevich, Usp. Fiz. Nauk **155**, 3 (1988) [Sov. Phys. Usp. **31**, 385 (1988)].

16. K. M. Salikhov, Yu. N. Molin, R. Z. Sagdeev, and A. L. Buchachenko, *Spin Polarization and Magnetic Field Effects in Radical Reactions*, Ed. by Yu. N. Molin (Elsevier, Amsterdam, 1984).

17. Yu. A. Osip'yan, S. I. Bredikhin, V. V. Kveder, *et al.*, *Electronic Properties of Dislocations in Semiconductors* (Editorial URSS, Moscow, 2000), p. 46 [in Russian].

18. A. L. Buchachenko, *Spin Chemistry* (Mosk. Gos. Univ., Moscow, 2002) [in Russian].

19. A. L. Buchachenko, Usp. Khim. **62**, 1139 (1993).

20. P. A. Schnegg, C. Jaccard, and M. Aegerter, Phys. Status Solidi B **63**, 587 (1974).

21. *Molecular Magnetism (New Magnetic Materials)*, Ed. by K. Itoh and M. Kinoshita (Gordon and Breach, Kodansha, Tokyo, 2000), p. 75.

22. E. J. Sharp and D. A. Avery, Phys. Rev. **158**, 511 (1967).

23. Yu. A. Ossipyan, R. B. Morgunov, A. A. Baskakov, *et al.*, Phys. Status Solidi A **201**, 148 (2004).

24. M. Górska, J. R. Anderson, G. Kido, *et al.*, Phys. Rev. B **45**, 11702 (1992).

25. R. B. Morgunov, S. Z. Shmurak, A. A. Baskakov, and Y. Tanimoto, Zh. Éksp. Teor. Fiz. **124**, 840 (2003) [JETP **97**, 754 (2003)].

26. J. E. Srtutt and E. Lilley, Phys. Status Solidi A **33**, 229 (1976).

27. J. R. Anderson, G. Kido, Y. Nishina, *et al.*, Phys. Rev. B **41**, 1014 (1990).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 100, No. 1, 2005, pp. 77–88.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 1, 2005, pp. 87–99.

ORDER, DISORDER, AND PHASE TRANSITIONS IN CONDENSED SYSTEMS

Noise Pumping of Nuclear Spin Waves in an Antiferromagnet

A. V. Andrienko

Institute of Molecular Physics, Russian Research Center Kurchatov Institute, pl. Kurchatova 1, Moscow, 123182 Russia

e-mail: andrienko@imp.kiae.ru

Received July 1, 2004

Abstract—Parametric excitation and the above-threshold behavior of a nonequilibrium system of nuclear magnons are investigated both under microwave noise pumping and under monochromatic pumping with noise modulation of the magnon spectrum. It is established that there exist two critical pumping amplitudes: the first corresponds to the onset of nonlinear absorption of the microwave field, and the second corresponds to the development of strong phase correlations—a nonequilibrium Bose condensate (NBC)—in a system of excited magnon pairs. The formation of the NBC was recorded by two methods: by intense electromagnetic radiation from a sample and by a coherent response of parametric magnons to a harmonic modulation of their spectrum (by a modulation response). The existing theories provide a satisfactory description for the functional dependence of pumping thresholds on the parameters of experiment, except for the region of minimal temperatures. The observed discrepancy between theoretical and experimental results can be explained under the assumption that quasi-elastic two-magnon relaxation processes make a nonadditive contribution to the threshold amplitudes of noise pumping. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

A hyperfine interaction between the spins of the nuclei and the electron shells of magnetic ions in weakly anisotropic antiferromagnets has given rise to a new branch of coupled oscillations of electronic and nuclear spins—nuclear spin waves (NSWs). The frequencies of these oscillations lie in the range of NMR frequencies ($\omega_n/2\pi \le 700$ MHz for ^{55}Mn nuclei), and the wavevectors $\mathbf{k} \le 10^6$ cm^{-1}. The most remarkable feature of NSWs is that they exist in a paramagnetic system of nuclear spins I that is characterized by a high level of thermal fluctuations and weak polarization ($\langle I\rangle/I \sim 1\%$) at liquid helium temperatures. The concept of NSWs was proposed in [1], where the authors calculated the spectra of these waves in ferro- and antiferromagnets and showed that the most convenient objects for investigating NSWs are cubic antiferromagnets and antiferromagnets with easy-plane anisotropy. The characteristic feature of these antiferromagnets is the presence of a low-frequency branch of spin waves and the exchange enhancement of hyperfine interaction [2]. The NSW spectrum in the easy-plane antiferromagnet $CsMnF_3$ is given by

$$\omega_k = \omega_n[1 - (H_\Delta^2/(H^2 + H_\Delta^2 + \alpha^2 k^2))]^{1/2}, \qquad (1)$$

where H_Δ^2 [kOe^2] $= 6.4/T$ is the parameter of hyperfine interaction, T is temperature, H is a static magnetic field, $\alpha = 0.95 \times 10^{-5}$ kOe cm is the constant of inhomogeneous exchange interaction, and $\omega_n/2\pi = 666$ MHz.

Spin and magnetoelastic waves in magnets represent a very convenient object for investigating the physics of nonlinear wave processes. At present, parametric resonance of these waves has been studied sufficiently well in an ac magnetic field $\mathbf{h}\cos\omega_p t$ parallel to a dc field $\mathbf{H}$. When the strength of the microwave magnetic field exceeds a threshold value h_c, the sample exhibits parametric instability with respect to the decay of a pumping quantum into a pair of magnons (or quasiphonons) of half the frequency with equal and oppositely directed wavevectors ($\omega_p = \omega_\mathbf{k} + \omega_{-\mathbf{k}}$). The critical amplitude h_c of the microwave field, which is usually called a pumping threshold or a parametric-excitation threshold, is defined by the expression $h_c = \gamma/V$, where γ is the relaxation rate of the excited waves and $V = (\partial\omega_\mathbf{k}/\partial H)/2$ is the coefficient of coupling between parametric waves and the pumping field; this coefficient is determined by the effective magnetic moment of the excited wave. For NSWs, this coefficient is defined by

$$V = \frac{H}{2H_\Delta^2}\frac{(\omega_n^2 - \omega_k^2)^2}{\omega_n^2\omega_k}; \qquad (2)$$

i.e., it depends on the frequency of magnons, magnetic field, and temperature.

Immediately above the excitation threshold (i.e., for $h > h_c$), a dynamic order is established in the system,

which is characterized by two parameters: the number of parametric pairs of magnons (or quasiphonons) and their phase with respect to the pumping field (see [3, 4]). Obviously, this nonequilibrium Bose condensate (NBC) of a macroscopic number of generated pairs represents a forced oscillation of the medium at the frequency of the external field.

The situation is substantially complicated when the external ac field is applied in the frequency interval $(\omega_p - \Delta\omega/2, \omega_p + \Delta\omega/2)$ and the frequency bandwidth of pumping is greater than the γ of the excited waves. The very possibility of parametric excitation of an oscillator by a noise field seems to be rather questionable [5]. Moreover, the possibility of nonequilibrium phase transition accompanied by the formation of an NBC in the wave system excited by noise pumping is still more questionable, because a noise field does not have any distinguished phase and cannot establish a coherent state in the excited system by itself.

Parametric resonance of spin waves under noise pumping was considered by Mikhailov and Uporov [5] and Cherepanov [6]. These authors obtained identical (up to a factor of 2 for $\Delta\omega \gg \gamma$) formulas for the mean threshold power P_n of the noise field at which time-average nonlinear absorption starts in the system. Under Gaussian noise pumping for arbitrary $\Delta\omega \ll \omega_p$, the threshold power is given by [7]

$$(P_n/P_c) - 1 = \Delta\omega/\gamma, \qquad (3)$$

where $P_c \propto h_c^2$ is the threshold power for monochromatic pumping and the critical power P_n represents the integral power of microwave noise pumping (rather than its spectral density).

However, these authors have fundamentally different views on the above-threshold behavior of a system of nonequilibrium magnons. In [5], the authors argue that the basic limiting mechanism for the microwave absorption is nonlinear dissipation, while the phase correlations of waves are ineffective. Conversely, the author of [6] argues that strong phase correlations should arise in the pairs of excited waves (the phase limiting mechanism), like under monochromatic pumping.

Such phase correlations were experimentally observed for the first time by Andrienko and Safonov [7] under noise pumping of magnetoelastic waves in the antiferromagnetic $FeBO_3$; they observed two thresholds of instability. The first threshold (P_{n1}) corresponds to the onset of nonlinear absorption of microwaves, while the second (P_{n2}) corresponds to the formation of an NBC. (Note that the authors of [5, 6] did not consider the possibility of two thresholds.) However, because of the specific features of the experiment in [7], namely, strongly anisotropic excitation (quasiphonons are generated along the C_3 axis of the crystal) and a possible effect of the boundaries of a sample on the formation of an NBC (the thickness of the plane-parallel plate was much less than the mean free path of a magnetoelastic wave), the authors could not answer the question whether or not an NBC can be formed in an infinite sample. Finally, in [8], an NBC was first observed under noise pumping of nuclear magnons in the antiferromagnetic $CsMnF_3$ when the mean free path of magnons was less than the sample size and the effect of boundaries could be neglected.

Another possibility of the noncoherent effect of external fields on a spin-wave system was first investigated by Zautkin and colleagues [9]. They theoretically and experimentally investigated a parametric generation of electron magnons by a monochromatic microwave field in yttrium iron garnet (YIG) under a noise modulation of the spectrum of spin waves by a radio-frequency field and argued that this case corresponds to nonmonochromatic (noise) pumping with fluctuating phase. They calculated and measured the threshold pumping amplitude corresponding to the onset of nonlinear absorption and the nonlinear magnetic susceptibility of a sample. The threshold formula given in [9] is analogous to Eq. (3) and is expressed as

$$(P_n/P_c) - 1 = (VH_n)^2/\gamma = H_n^2(V/h_c), \qquad (4)$$

where H_n is the spectral density of the noise field; the threshold of "noise" pumping P_n in this case (in contrast to (3)) is the threshold power of monochromatic pumping under the noise modulation of the spectrum of spin waves. Later, Safonov [10] considered the decay threshold of phase correlations in a wave system excited by a monochromatic field when the noise modulation of the spectrum was switched on, and obtained the formula

$$h_n/h_c - 1 = (VH_n)^2/\gamma,$$

which can conveniently be rewritten as

$$(P_n/P_c) - 1 = H_n^2(V/h_c)[2 + H_n^2(V/h_c)]. \qquad (5)$$

For the present, this is actually the only calculation of the threshold P_{n2}, i.e., the destruction (formation) threshold for an NBC. In the limit of small H_n, the threshold powers P_{n1} and P_{n2}, defined by (4) and (5), respectively, differ only by a factor of 2.

In [9], based on the qualitative agreement between theory and experiment, the authors concluded that phase correlations are significant. However, they could not directly confirm the presence of strong phase correlations in a wave system under noise modulation of the magnon spectrum. Note that the experimental results of [9] can equally (up to the measurement errors of the pumping threshold P_n) be described by formulas (4) and (5). The formation of an NBC under parametric excitation of waves (quasiphonons) with the noise

modulation of their spectrum was first observed by Andrienko and Safonov in [7].

In the present paper, we investigate the behavior of the system of nuclear magnons both under noise pumping with $\gamma \leq \Delta\omega \ll \omega_p$ and under monochromatic microwave pumping with the noise modulation of the spectrum of NSWs. In this case, the mean free path of generated magnons is much less than the crystal size, whereby the effect of sample boundaries on the formation of a stationary state of parametric pairs can be neglected. Moreover, under the conditions of coherent pumping and noise pumping, we investigate the radiation of electromagnetic waves by parametric pairs of NSWs. This effect has been first discovered in the present study.

2. EXPERIMENTAL TECHNIQUE

Nuclear magnons in the easy-plane antiferromagnet $CsMnF_3$ (with the Néel temperature 53.5 K) were generated by a microwave field of frequency $\omega_p/2\pi$ = 1080–1180 MHz. The noise field was generated by two methods.

(1) Noise modulation. In this case, a microwave oscillator operated in the usual mode (monochromatic pumping) and a signal from a white noise generator operating in the frequency range from 0.015 to 600 kHz was amplified by a broadband amplifier and was applied to a modulating coil that was wound coaxially with a helical microwave resonator. The ripple amplitude of the noise spectral density was no greater than 3 dB. The measurement accuracy of the mean spectral density H_n of the noise field was 15%, and the accuracy of relative measurements was 5%. To a high degree of accuracy, such noise can be considered as white noise because its bandwidth is much greater than the relaxation frequency of magnons $\gamma/2\pi \leq 20$ kHz.

(2) Noise pumping. A noise signal from a radio-frequency white noise generator was used for modulating the frequency of a microwave oscillator. The spectrum of the output microwave-pumping signal was investigated by a spectrum analyzer. The shape of the spectrum was close to a Gaussian curve with a frequency bandwidth of $\Delta\omega \ll \omega_p$. The bandwidth of the pumping spectrum was measured at a half maximum to an accuracy of 10%, the condition $\gamma < \Delta\omega < \Delta\omega_R$ being always satisfied, where $\Delta\omega_R$ is the linewidth of the resonance curve of the microwave resonator loaded with a sample.

A single-crystal sample of $CsMnF_3$ with a size of about $3 \times 3 \times 5$ mm^3 was placed in a helical microwave resonator with a quality factor of $Q \sim 500$ filled with liquid helium. The experiment was carried out in the temperature interval T = 1.9–4.2 K in a magnetic field of H = 500–2000 Oe under parallel microwave pumping (**h** || **H**). The relaxation rate $\gamma/2\pi$ of generated magnons was calculated using the monochromatic pumping threshold h_c with an accuracy of 25% and ranged (depending on T and H) from 6 to 20 kHz. The accuracy of relative measurements of h_c and γ was 5%. The mean free path of magnons was $L \leq 1$ mm, i.e., several times less than the characteristic size of the sample.

Microwave pumping was performed in continuous-wave mode, as well as in a pulsed mode with a repetition rate of 10–100 Hz and a pulse length of 300–2000 μs. In the pulsed mode, the threshold pumping power was determined by the appearance of a nonlinear distortion (chip) on a detected microwave pulse transmitted through the resonator. In continuous-wave mode, this threshold is determined by the appearance of a modulation response in the spectrum of the microwave signal transmitted through the resonator loaded with a sample (the description of this technique is given below). The relative measurement accuracy of the threshold power was 10%.

We used two methods to detect an NBC, which are based on the observation of collective phenomena associated with the formation of this state.

The first method consist in the observation of the modulation response α_m. This method was developed and widely used for investigating the properties of the NBC of magnons that is formed under monochromatic pumping [4]. Later, this method was successfully applied to detect the NBC of quasiphonons generated by noise pumping [7]. A weak radio-frequency magnetic field $H_m\cos(\omega_m t)$ is applied to a sample to modulate the spectrum of spin waves. This gives rise to oscillations of the amplitude and phase of the NBC about their equilibrium values (the so-called collective oscillations), which result in a modulation, with amplitude ΔP and frequency ω_m, of the microwave power absorbed by the sample. This amplitude modulation ($\Delta P = \alpha_m H_m$) points to the existence of the NBC in the sample.

To record α_m, a microwave signal transmitted through the resonator was detected and applied to a spectrum analyzer or to a selective microvoltmeter tuned to the frequency ω_m and then fed to the Y input of an X–Y automatic recorder.

The second method consists in the observation of electromagnetic radiation from a system of excited nuclear magnons. Earlier, this effect was observed in a system of magnetoelastic waves [10]. It consists in the fact that a sample in which an NBC is produced by monochromatic microwave pumping emits characteristic electromagnetic radiation after switching off the pumping. The frequency of this radiation is close to the pumping frequency, and its intensity is nonmonotonic in time and has a magnitude comparable to the microwave power absorbed by the sample. In the present paper, we could also observe similar radiation from a system of NSWs.

3. RESULTS AND DISCUSSION

3.1. Parallel-Pumping Thresholds under Noise Modulation of the Magnon Spectrum

We observed two pumping thresholds $P_{n1} < P_{n2}$. The first threshold corresponds to the onset of nonlinear absorption of microwave power, and the second corresponds to the formation of an NBC. Since the difference between P_{n1} and P_{n2} is usually no greater than the experimental error, we only traced the behavior of the threshold P_{n1} in our experiments, because it is easier to measure. In this section, we denote it simply by P_n. Figure 1 shows the threshold power of monochromatic parallel pumping as a function of the spectral density of the noise field for various temperatures, pumping frequencies, and magnetic fields. One can see that $(P_n/P_c - 1)$ is proportional to the squared spectral density of noise, depends on H and T, and rapidly decreases as the frequency of generated magnons approaches ω_n. The theoretical curves calculated by formula (4) with a fitting factor of 5.9 (for $\omega_p/2\pi \approx 1090$ MHz) are shown by a solid line. They well describe the functional dependence of the threshold on H_n^2 and T at this pumping frequency over the entire range of H_n. The theoretical curve calculated by (5) (dashed curves) describes the functional dependence of the threshold only for $P_n/P_c < 5$; but the fitting parameter for this curve is less; it is just 2.5. As the pumping frequency increases, the effect of H_n on the threshold should rapidly decrease because, according to (2), the parameter V that enters (4) rapidly decreases. This decrease is clearly demonstrated in Fig. 1, where the upper (○) and lower (•) rows of dots differ only by the pumping frequency. Moreover, this decrease turns out to be even stronger than that predicted by the theory, so that the fitting parameters decrease and, for $\omega_p/2\pi \approx 1166$ MHz, are 3.4 and 1.6, respectively. Recall that, actually, the theory does not contain any fitting parameters, and the calculated value of P_n depends only on the accuracy of the absolute measurement of the threshold value of h_c and the spectral density H_n of the noise field. Actually, the calculations are performed for ideal white noise, while the real frequency characteristic $H_n(\omega_m)$ of the noise-field generator is not uniform and has a limited bandwidth of 0.15–600 kHz. Therefore, a calculated threshold may not coincide with a measured one; however, this may barely at all affect the determination of the absolute value and the functional dependence of the pumping threshold. Thus, the effect of the noise modulation of the spectrum on the pumping threshold for NSWs turns out to be several times greater than the calculated result.

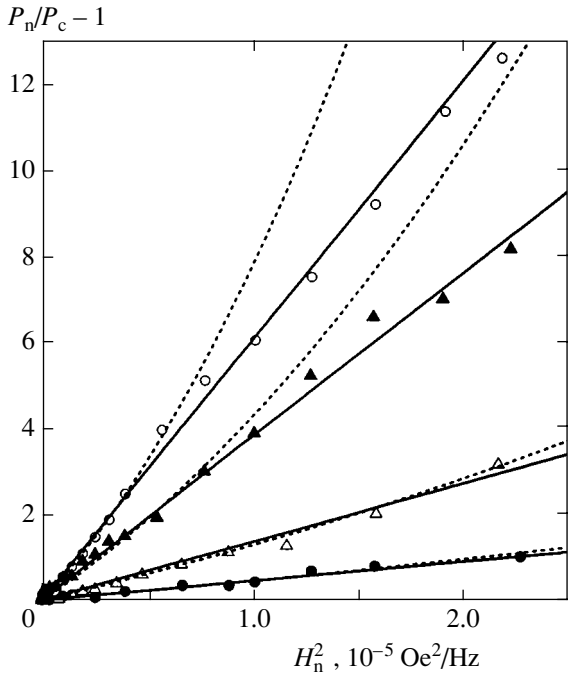


Fig. 1. Relative increase in the threshold power of coherent pumping as a function of the spectral density of noise field; (○) H = 1830 Oe, T = 2.05 K, and $\omega_p/2\pi$ = 1088.8 MHz; (▲) H = 1134 Oe, T = 4.2 K, and $\omega_p/2\pi$ = 1093.3 MHz; (△) H = 689 Oe, T = 4.2 K, and $\omega_p/2\pi$ = 1093.3 MHz; and (●) H = 1775 Oe, T = 2.0 K, and $\omega_p/2\pi$ = 1165.8 MHz. Solid curves correspond to theoretical calculations by formula (4), and dashed curves, to calculations by formula (5). Fitting factors are equal to 5.9 and 2.5 for $\omega_p/2\pi$ = 1090 MHz and 3.4 and 1.6 for $\omega_p/2\pi$ = 1166 MHz.

The field dependence of the threshold of monochromatic pumping at a temperature of T = 2.08 K for fixed values of the amplitude of noise modulation are shown in Fig. 2. Figure 3 represents the same results as $(P_n/P_c - 1)$ vs. H together with the similar results obtained at T = 4.2 K. One can see that the fitting parameter is independent of H and T; i.e., the theoretical calculations satisfactorily describe the field and temperature dependence of the parallel-pumping thresholds for NSWs under the noise modulation of the spectrum. Thus, at a fixed pumping frequency, the theory describes the functional dependence of the threshold on any parameter: on the spectral density of noise, on a magnetic field, and on temperature. The NSW parameters, such as relaxation rate, wavevector, and the coupling coefficient between NSWs and pumping, vary several times in these intervals of H and T. Hence, the fitting factor is independent of the above parameters of magnons; it depends only on their frequency. Thus, the measurements show that, although the theory gives an incorrect absolute value of the threshold, it fails to describe only one function of the threshold, namely, the frequency dependence of the monochromatic pumping threshold under the noise modulation of the spectrum of nuclear magnons. Note that the effect of the excitation frequency on the noise pumping threshold is studied here for the first time. Earlier, the values of the pumping threshold measured during the study of nuclear magnons in YIG [9] under the noise modulation of the spectrum coincided with the calculated ones up to the measurement errors. However, these measure-

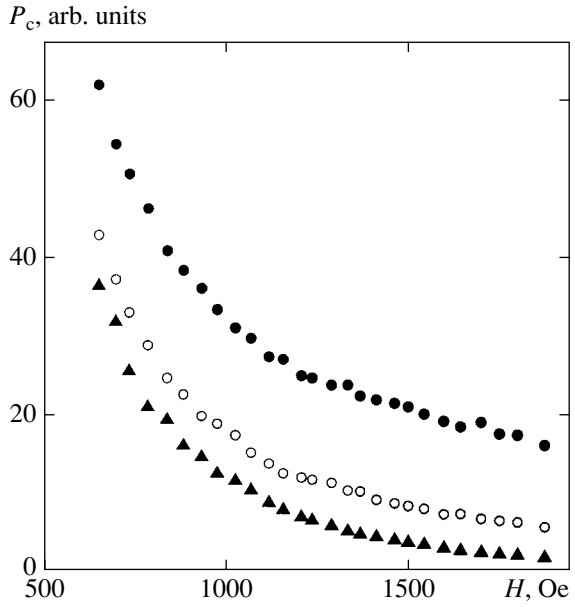


Fig. 2. Threshold power of coherent pumping as a function of a static magnetic field for a fixed amplitude of noise modulation of the magnon spectrum: (▲) $H_n = 0$, (○) $H_n = 0.02$ Oe/Hz$^{1/2}$, and (●) $H_n = 0.04$ Oe/Hz$^{1/2}$; $T = 2.08$ K and $\omega_p/2\pi = 1094.4$ MHz.

ments were carried out at a single value of pumping frequency and only for two values of the magnetic field. In the experiments on the excitation of quasiphonons in $FeBO_3$ [7], the theory failed to describe the dependence of P_n on the magnetic field.

3.2. Noise Pumping Thresholds of NSWs

Investigations have shown that one should also distinguish two thresholds $P_{n1} < P_{n2}$ under noise pumping of nuclear magnons. The first threshold corresponds to the onset of nonlinear absorption of microwave power, and the second, to the formation of an NBC. The ratio P_{n2}/P_{n1} of these thresholds increases with $\Delta\omega$ but does not exceed 1.25. Figure 4 shows the threshold of parametric pumping of NSWs as a function of the frequency bandwidth of microwave pumping. Since the difference between the thresholds P_{n1} and P_{n2} is usually no greater than the experimental error, we only kept track of P_{n1} in these experiments; in what follows, we simply denote it by P_n. According to theory (3), $(P_n/P_c - 1)$ is proportional to $\Delta\omega$ for any values of T, H, and ω_p. However, for a fixed pumping frequency of about 1090 MHz, the calculated value of $(P_n/P_c - 1)$ exceeds the measured value by a factor of 3 at temperature $T = 4.2$ K and by a

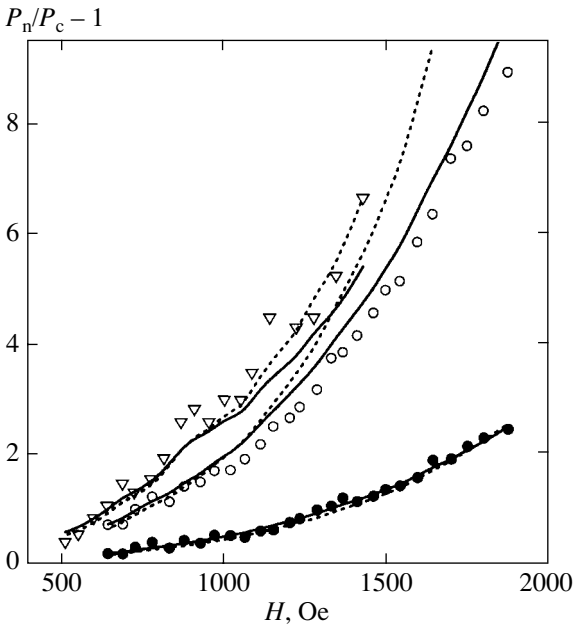


Fig. 3. Relative increase in the threshold power of coherent pumping as a function of a static magnetic field for a fixed amplitude of noise modulation; (●) $H_n = 0.02$ Oe/Hz$^{1/2}$, $T = 2.08$ K, and $\omega_p/2\pi = 1094.4$ MHz; (○) $H_n = 0.04$ Oe/Hz$^{1/2}$, $T = 2.08$ K, and $\omega_p/2\pi = 1094.4$ MHz; and (▽) $H_n = 0.03$ Oe/Hz$^{1/2}$, $T = 4.2$ K, and $\omega_p/2\pi = 1093.3$ MHz. Solid curves correspond to theoretical calculations by formula (4) with a fitting parameter of 5.9, and dashed curves, to calculations by formula (5) with a fitting parameter of 2.5.

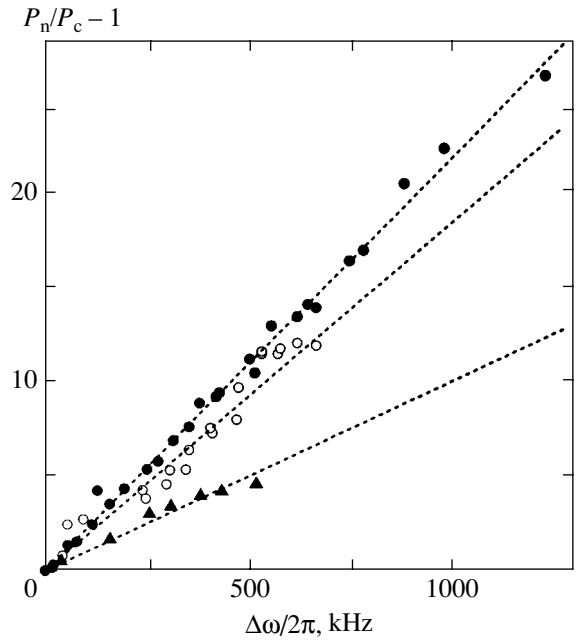


Fig. 4. Relative increase in the threshold power as a function of the frequency bandwidth of noise pumping; (●) $T = 2.05$ K, $\omega_p/2\pi = 1088$ MHz, $H = 1750$ Oe, and $\gamma/2\pi = 7.1$ kHz; (○) $T = 4.2$ K, $\omega_p/2\pi = 1093.3$ MHz, $H = 1680$ Oe, and $\gamma/2\pi = 16.8$ kHz; and (▲) $T = 2.0$ K, $\omega_p/2\pi = 1165.8$ MHz, $H = 1825$ Oe, and $\gamma/2\pi = 10.1$ kHz. Dashed curves correspond to calculations by formula (3) with the fitting parameters 0.15, 0.31, and 0.1 for the dots ●, ○, and ▲, respectively.

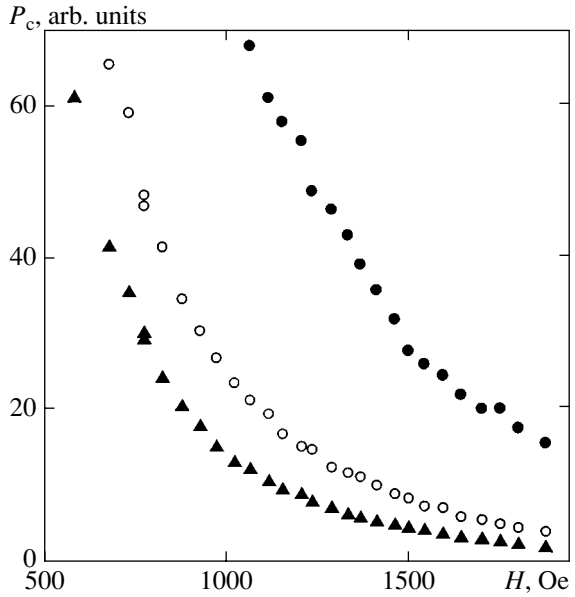


Fig. 5. Threshold power of noise pumping as a function of a static magnetic field for fixed values of the pumping bandwidth; (▲) $\Delta\omega/2\pi \approx 0$, (○) $\Delta\omega/2\pi = 40$ kHz, and (●) $\Delta\omega/2\pi = 270$ kHz; $T = 2.08$ K and $\omega_p/2\pi = 1094.4$ MHz.

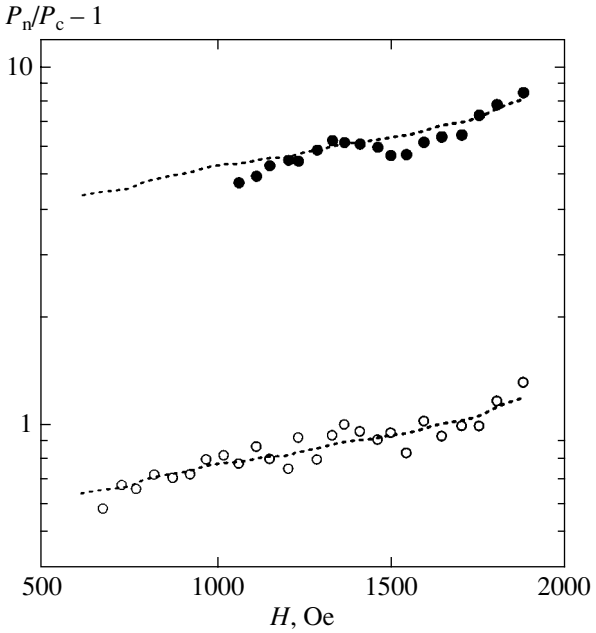


Fig. 6. Relative increase in the threshold power of noise pumping as a function of a static magnetic field for fixed values of the pumping bandwidth; (○) $\Delta\omega/2\pi = 40$ kHz and (●) $\Delta\omega/2\pi = 270$ kHz; $T = 2.08$ K and $\omega_p/2\pi = 1094.4$ MHz. The curves correspond to calculations by formula (3) with a fitting parameter of 0.175.

factor of about 7 at temperature $T = 2.05$ K. The fitting factor is also changed under the variation of the pumping frequency: at 1166 MHz, the measured value of $P_n/P_c - 1$ is less than the theoretical value by a factor of 10. Thus, in the case of noise pumping, the theory describes neither frequency nor temperature dependence of the generation threshold.

Figure 5 shows measured values of the noise pumping threshold as a function of a static magnetic field for fixed values of $\Delta\omega$. Figure 6 reproduces the same results as $P_n/P_c - 1$ vs. H. The calculated curves are obtained as follows. Using the measured monochromatic pumping threshold $h_c(H)$, we calculated $\gamma(H)$ by formula (2) and substituted it into (3) to calculate H as a function of $P_n/P_c - 1$. Again, one can see that the theoretical results with the above-mentioned fitting parameters are in good qualitative agreement with the experimental results. Here, it is essential that P_c varies 40 times over the range of magnetic fields considered. The wavevector, the relaxation rate of generated nuclear magnons, and the coupling coefficients between pumping and NSWs are also changed significantly. However, the fitting factor remains constant. Hence, the theory correctly describes the noise pumping threshold P_n as a function of the above-mentioned parameters of generated magnons, but gives a value of the threshold power that is several times greater than the measured value. Recall that, conversely, under the noise modulation of the spectrum of NSWs, the theoretical results were several times less than the experimental data. The field dependence of the noise pumping threshold obtained earlier when investigating the generation of quasiphonons in $FeBO_3$ [7] did not admit theoretical description. Thus, formula (3) much better describes the noise pumping of NSWs. This is likely to be associated with a weak effect of the sample boundaries on the system of generated magnons; i.e., for NSWs, the theoretical model is closer to the real experimental situation.

Under noise pumping of NSWs, theoretical results significantly deviate from experimental results as temperature decreases and frequency varies. Recall that the effect of the microwave-field frequency on the noise pumping is investigated in the present paper for the first time; all measurements in [7–9] were carried out at a fixed frequency.

One may propose the following qualitative explanation for this discrepancy between theory and experiment.

The analysis of the relaxation mechanisms of NSWs has shown [2] that, for $\omega_p/2\pi \approx 1090$ MHz and $T = 4.2$ K, the relaxation of NSWs in $CsMnF_3$ is determined, by about 50%, by inelastic scattering of NSWs by the fluctuations of nuclear magnetization (by two-magnon processes) and, by another 50%, it is determined by three-particle relaxation processes with the temperature dependence T^5. At 2.2 K, the contribution of inelastic relaxation processes decreases to about 5%. Moreover, the efficiency of three-particle processes rapidly decreases as the magnon frequency, i.e., the pumping frequency, increases. As a result, elastic pro-

cesses also become dominant. In our opinion, it is the dominant role of elastic relaxation processes that is responsible for the fact that the noise pumping threshold of NSWs proves to be appreciably less than the calculated one; this is especially manifest as temperature decreases or as the pumping frequency increases.

An assumption about the specific role of two-magnon processes and their nonadditive contribution to the threshold formula has been made by Zakharov and L'vov [11]. They suggested that, since the wave energy is not changed under elastic scattering (only the propagation direction of a wave is changed), secondary magnons keep taking part in the process of pumping, thus reducing the threshold of parametric excitation, which is described in this case by

$$h_c = [\gamma_{nel}(\gamma_{nel} + \gamma_{el})]^{1/2}/V, \quad (6)$$

where γ_{el} and γ_{nel} are the relaxation rates of elastic and inelastic processes, respectively. However, experiments with monochromatic pumping [2, 12] did not confirm this assumption. Experiments showed an additive contribution of all γ to the threshold amplitude: $h_c = (\gamma_{nel} + \gamma_{el})/V$. Apparently, the point is that, a crucial role under monochromatic pumping is played by the phase of a wave, which changes after a quasi-elastic scattering of a magnon. Moreover, quasi-elastic scattering may be accompanied by a slight variation in the energy and the absolute value of the wavevector of a magnon (the so-called drift over the spectrum). As a result, a secondary magnon is out of resonance with the pumping field. In the case of microwave noise pumping, whose frequency and phase rapidly fluctuate, neither the variation of the phase of secondary magnons nor the weak drift of magnons over the spectrum of NSWs (within $\Delta\omega$) influence the coupling between magnons and the driving microwave field. Therefore, secondary magnons continue to take part in the pumping process, thus reducing its threshold. (These secondary magnons can also be interpreted as an additional noise field.) Thus, it is the noise pumping under which the contribution of two-magnon processes to the threshold amplitude should become nonadditive. Our experiments have shown that the greater the relative contribution of elastic processes, the greater the deviation of the fitting parameter from 1, displayed both by temperature and frequency dependence of the fitting parameter. Unfortunately, the noise pumping threshold with regard to the specific role of quasi-elastic processes has not been calculated. Nevertheless, the qualitative model proposed above is in good agreement with experiment.

3.3. Kinetics of Parametric Instability

It is well known [3] that, in the case of monochromatic pumping, when the amplitude of the microwave field exceeds a certain threshold value h_c, the number of magnons first increases by the exponential law

$$N_k = N_0 \exp[(h/h_c - 1)(t/\tau_m)], \quad (7)$$

where N_0 is on the order of the number of thermal magnons of frequency ω_k, τ_m is the lifetime of these magnons, and t is the time measured from the switching moment of a microwave pumping pulse. The number of magnons increases due to the sample's absorption of photons stored in the resonator. This absorption also increases exponentially at first. When the number of magnons reaches N_{k1} (which is determined by the sensitivity of the setup), a chip appears on the microwave pulse transmitted through the resonator. The time interval $\tau = \tau(N_{k1})$ from the beginning of the pulse to the chip is given by the expression

$$1/\tau = C(h/h_c - 1), \quad (8)$$

where C is a constant determined by the lifetime of magnons and the sensitivity of the experimental setup. In a certain period of time after the chip, due to the restriction mechanisms, the number of parametric pairs and the magnitude of microwave absorption reach a steady-state level. For monochromatic pumping, the linear dependence of the inverse time to the chip on the supercriticality was repeatedly verified in experiments (see, for example, [3, 13]).

In this section, we focus on the development of parametric instability under noise pumping, in particular, on the dependence of the time τ on the spectral bandwidth of the pumping $\Delta\omega$ and on the noise modulation amplitude of the magnon spectrum H_n.

The development time of parametric instability as a function of noise level was measured as follows. For fixed values of the microwave power and pumping bandwidth (P, $\Delta\omega$ = const), we increased the pulse length until a chip appears on it. This time τ to a chip slightly varied from pulse to pulse, the variation being the greater, the greater the pumping bandwidth $\Delta\omega$. We measured the minimal, rather than the mean, value of τ at which a chip appeared on microwave pulses. The results of these measurements are represented in Fig. 7.

The most obvious factor responsible for increasing the time to a chip with increasing the pumping bandwidth is a decrease in the supercriticality as $\Delta\omega$ increases. To take into account this change in the supercriticality, we should replace (h/h_c) by (h/h_n) in (8), where $h_n = h_n(\Delta\omega) \propto (P_n(\Delta\omega))^{1/2}$ is a function of the frequency bandwidth of pumping. Then, formula (8) is represented as $1/\tau = C[(P/P_n)^{1/2} - 1]$. Taking into account the experimental dependence of P_n on $\Delta\omega$ (which is described by theory (3) with a certain fitting factor b), we can rewrite this formula as

$$1/\tau = C[(h/h_c)/(b\Delta\omega/\gamma + 1)^{0.5} - 1], \quad (9)$$

where b is taken from earlier experiments and is equal to $b = 0.15$ for $\omega_p/2\pi = 1094.4$ MHz, $T = 2.08$ K, and

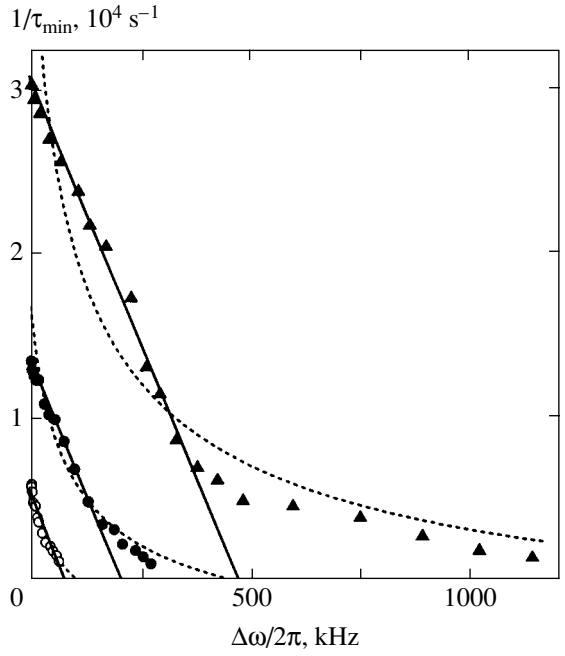


Fig. 7. Time to a chip as a function of the bandwidth of noise pumping for fixed values of pumping power; (○) P/P_c = 3.5, (●) P/P_c = 11.7, and (▲) P/P_c = 50.5; T = 2.08 K, H = 1870 Oe, and $\omega_p/2\pi$ = 1094.4 MHz. Dashed curves are calculated by formula (9), and the straight lines, by formula (10).

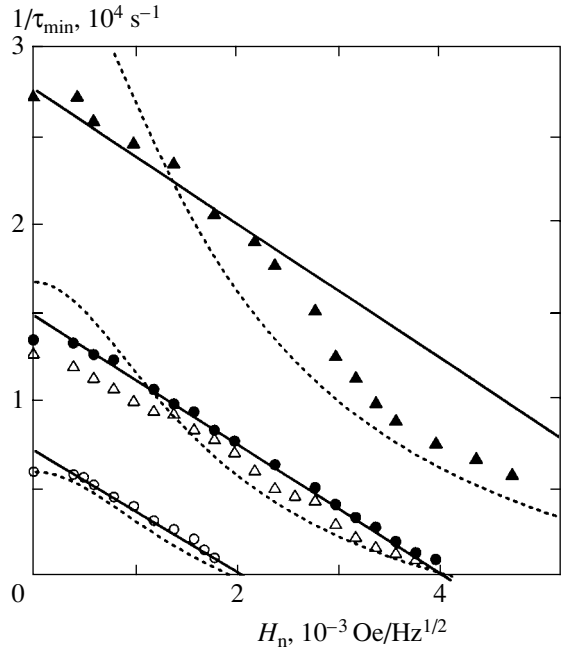


Fig. 8. Time to a chip of coherent pumping as a function of the noise modulation amplitude of the spectrum for several values of pumping power; (○) P/P_c = 3.4, (△) P/P_c = 10, (●) P/P_c = 11.4, and (▲) P/P_c = 38; T = 2.08 K, H = 1870 Oe, and $\omega_p/2\pi$ = 1094.4 MHz. Dashed curves are calculated by formula (11), and the straight lines, by formula (12).

H = 1870 Oe. Figure 7 shows that formula (9) with a fitting parameter of C = 7000 well describes the general behavior of the experimental curves, which is rather unexpected because formula (9) takes into account only the decrease in the supercriticality with increasing the pumping frequency bandwidth $\Delta\omega$ but does not take into account the specific feature of the noise-induced generation of parametric magnons. Note that, in a wide range of $\Delta\omega$, the experimental results for three values of supercriticality are well described by three parallel lines, which correspond to the empirical function

$$1/\tau = 1/\tau_0 - 10.3\Delta\omega, \qquad (10)$$

where τ_0 is the time to a chip under coherent pumping.

Similar measurements of the development time of parametric instability were also carried out under coherent pumping of NSWs with the noise modulation of the spectrum. The results of these measurements are shown in Fig. 8. The measurement technique is analogous: for constant values of the microwave power and noise modulation level (P, H_n = const), we increased the length of a pulse until a chip appears on it. Since the noise modulation increases the pumping threshold, to a first approximation, we can take into account the effect of this modulation on the development time of parametric instability by substituting the experimental dependence of the threshold on H_n into formula (8). As shown above, this dependence is well described by theory (4) with a fitting factor of 5.9. Using formulas (4) and (8) and the experimental values of $P_n(H_n)$, we obtain

$$1/\tau = C[(h/h_c)/(1 + 5.9H_n^2V/h_c)^{0.5} - 1]. \qquad (11)$$

The curves corresponding to (11) are shown in Fig. 8 by dashed lines. Since formula (11) does not take into account a shift in the phase of the NBC under noise modulation of the spectrum, the agreement between experiment and calculations by (11) can be assumed to be quite satisfactory. Incidentally, the only fitting parameter C in (11) is the same as in (9): C = 7000. It allows one to satisfactorily describe, by formulas (9) and (11), all experimental results represented in Figs. 7 and 8. Note that, for $P/P_c < 20$, the experimental results shown in Fig. 8 are also well described by straight lines drawn by the empirical formula

$$1/\tau = 1/\tau_0 - 3.37 \times 10^6 H_n. \qquad (12)$$

A certain similarity between empirical formulas (10) and (12) seems to be quite unexpected because the effect of H_n on the threshold power P_n is quadratic, whereas the effect of the bandwidth of the noise-pumping spectrum $\Delta\omega$ on P_n is linear.

3.4. Observation of Nonequilibrium Bose Condensation of Parametric Pairs of Magnons Generated by Noise Pumping

Up to now, we have dealt with the threshold amplitude of pumping above which a large number of nonequilibrium magnons are generated in a sample, and with the development time of parametric instability. Of greatest interest is the state of these nonequilibrium magnons (parametric pairs of magnons) that is formed above the threshold of noise pumping, in particular, the solution of the problem on the possibility of strong phase correlations in such a system.

Figure 9 represents the amplitudes of the modulation response α_m, drawn by an automatic recorder, as a function of the integral power of noise pumping and monochromatic pumping. These data are recorded in a continuous-wave mode. This figure also shows the oscillograms of microwave pulses transmitted through a resonator and recorded at three values of the amplitude of the microwave noise field. We observed two pumping thresholds, $P_{n1} < P_{n2}$. At point *A*, where $P < P_{n1}$, there is no nonlinear absorption, and $\alpha_m = 0$. Oscillogram *B* recorded above the threshold P_{n1} ($P_{n2} > P > P_{n1}$) shows "bursts" of absorption; i.e., there exist (on the average) nonequilibrium magnons in the sample, but $\alpha_m = 0$. This nonlinear absorption increases with the pumping power (the bursts occur more frequently). Above the threshold P_{n2}, the pulse has a shape of *C* in the figure, and a modulation response arises ($\alpha_m \neq 0$), which indicates that there exist (on the average) phase correlations in the system, i.e., that an NBC is formed. As we pointed out above, the ratio of the thresholds P_{n2}/P_{n1} increases with $\Delta\omega$ but does not exceed 1.25. This result is radically different from the case of noise pumping of quasiphonons [7], where the ratio of these two threshold powers amounted to 4. Such a small difference between the thresholds P_{n1} and P_{n2} is likely to be attributed to the dominant contribution of elastic processes to the relaxation of NSWs. It is obvious that the threshold of the NBC formation, P_{n2}, is associated with the total number of magnons within the pumping bandwidth $\Delta\omega$. When $P > P_{n1}$, nonequilibrium magnons generated by pumping arise in the sample. Secondary magnons that result from two-magnon scattering processes remain in the same spectral domain of NSWs. Hence, these magnons are accumulated; i.e., the occupation numbers of magnons with frequencies $\omega_k \approx \omega_p/2$ rapidly grow. As a result, the number of magnons reaches a critical value necessary for the formation of NBC earlier, for lower supercriticality (P/P_{n1}), compared with the case when three-particle relaxation processes (which decrease the occupation numbers of magnons near the frequency $\omega_p/2$) are dominant.

The existence of the NBC was also detected by electromagnetic radiation from a sample, which was observed immediately after switching off a pumping pulse (see Fig. 10). It is well-known that the power of a microwave signal transmitted through a resonator is proportional to the number of photons in the resonator. In the absence of the NBC, immediately after switching off the oscillator, this signal monotonically decreases

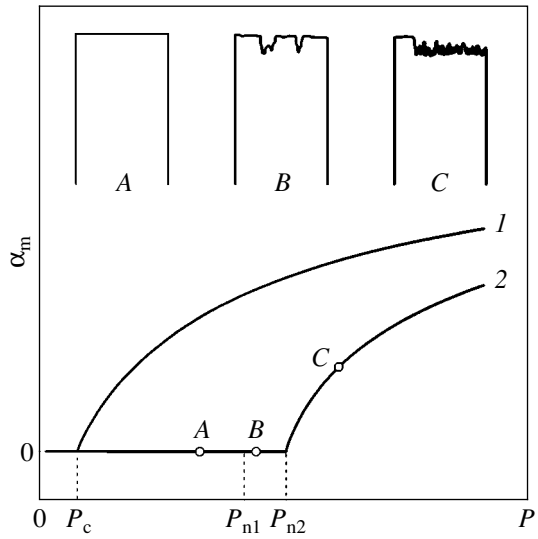


Fig. 9. Signal of modulation response as a function of power for (*1*) monochromatic pumping and (*2*) noise pumping (drawn by automatic recorder), and the oscillograms of microwave pulses of noise pumping transmitted through the resonator at points *A*, *B*, and *C*.

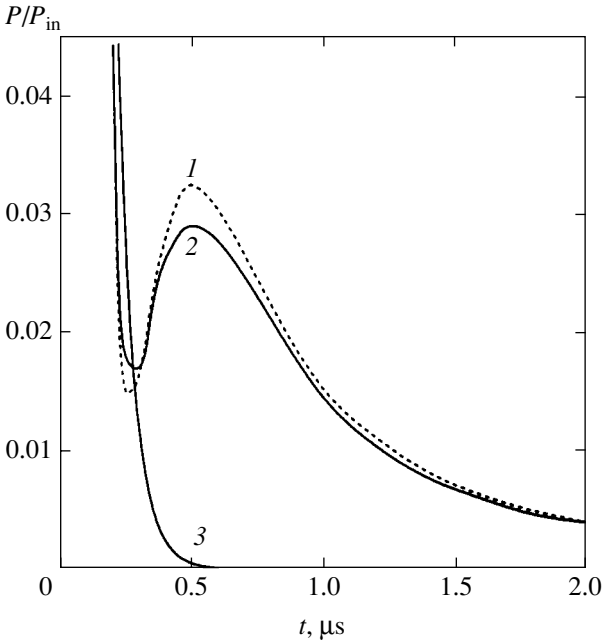


Fig. 10. Oscillograms of the trailing edge of microwave pulses and a signal of electromagnetic radiation from a sample for $T = 2.08$ K, $H = 1790$ Oe, $P_{in} \approx 2000P_c$, and $\omega_p = \omega_R - 2\pi \times 0.7$ MHz for (*1*) monochromatic pumping and (*2*) noise pumping. Curve (*3*) represents the trailing edge of a pulse for $P < P_c$; $t = 0$ corresponds to the moment of switching off the pump oscillator.

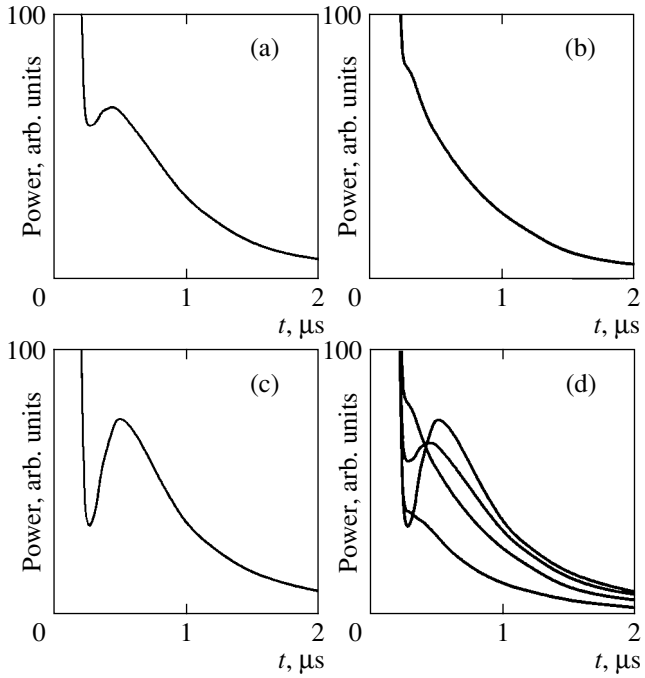


Fig. 11. Oscillograms of the trailing edge of microwave pulses and a signal of electromagnetic radiation from a sample for various shifts of the monochromatic pumping frequency with respect to the resonator frequency $\omega_R = 2\pi \times 1094.8$ MHz: (a) $\omega_p = \omega_R$, (b) $\omega_p = \omega_R + 2\pi \times 0.33$ MHz, (c) $\omega_p = \omega_R - 2\pi \times 0.74$ MHz, and (d) under noise pumping with $\omega_p = \omega_R - 2\pi \times 0.2$ MHz and $\Delta\omega = 2\pi \times 1$ MHz; $T = 2.08$ K, $H = 1840$ Oe, and $P \approx 2500P_c$.

with a characteristic decay time of photons of about 0.1 μs, which is determined by the Q of the resonator. If the pumping power is substantially greater than the threshold power, then, after switching off the pumping field, one can observe radiation from the sample behind the trailing edge of the pulse. First, the microwave power transmitted through the resonator sharply decreases (the trailing edge of the pulse decreases more abruptly than in the case $P < P_c$). Then, the signal starts to grow, and its amplitude reaches a maximum in $\tau \sim 0.5$ μs after the end of the pulse. Then, one can observe a decay of radiation with a characteristic time of about 1 μs. The maximal radiation power is comparable with the power of the pumping signal absorbed by the sample.

This radiation was earlier investigated in [14] under parametric excitation of magnetoelastic waves in $FeBO_3$. It is associated with the fusion of two quasiparticles (magnons in our case), which constitute a parametric pair, and is accompanied by the generation of a photon with the frequency close to ω_p ($m + m \longrightarrow ph$), i.e., by a process inverse to the pumping. Since parametric pairs in the NBC have equal phases, this radiation is coherent and has a large amplitude, in contrast to the radiation from individual magnons at frequency ω_k. Since the electromagnetic radiation from parametric pairs of NSWs is observed for the first time in the present study and because it provides information on the properties of a Bose condensate of magnons generated by noise pumping, we will dwell on this phenomenon in greater detail.

In [15], Safonov has shown that a new mechanism of positive nonlinear damping of parametric pairs is possible under parallel pumping in the combined resonator–sample system—a radiation damping due to the emission of a photon of frequency $2\omega_k$ by a parametric pair. The magnitude of the radiation damping is given by [14]

$$\gamma_{rad} = 2\pi\hbar Q V^2 N_k / v_R, \qquad (13)$$

where Q and v_R are the Q factor and the volume of the resonator, respectively. It is obvious that the emission of photons due to this process is hard to detect under switched on pumping because a counterflow of photons from the sample to the resonator merely reduces the microwave power absorbed by the sample. However, this radiation can be detected after the termination of the pumping pulse. It seems that such radiation was observed for the first time in a system of parametrically generated electronic magnons in YIG ferrite [16]; however, detailed information about this effect is not available. Later, this radiation was investigated in [14, 17] under parametric excitation of magnetoelastic waves in $FeBO_3$.

First, we consider electromagnetic radiation observed under monochromatic pumping of NSWs. It is important that the shape of the radiation signal depends both on the pumping power and on the relation between the natural frequency ω_R of the microwave resonator and the pumping frequency. While the pumping power affects only the amplitude (power) of radiation, the variation in the pumping frequency strongly affects the shape of the radiation signal as well. This is associated with the fact that, irrespective of the pumping frequency, the frequency of the damped signal of the resonator (curve *3* in Fig. 10) is always close to ω_R. Radiation from the sample is emitted at the natural frequency of the NBC, which is close to the pumping frequency: $\omega_{NBC} = \omega_p + \Omega$, where $\Omega \sim \gamma[P/P_c - 1]^{1/2}$ is the frequency of collective oscillations in the NBC system. The nonmonotonic behavior of radiation is attributed to the beating in the system of these two damped oscillators.[1] Hence, the shape of the radiation signal depends on the relation between the frequencies of these oscillators. Curve *1* in Fig. 10 is obtained for $\omega_p = \omega_R - 2\pi \times 0.7$ MHz, when the radiation peak is most clearly displayed. Shifting the pumping frequency with respect to the resonator frequency, one can observe radiation signals of various shapes. Figs. 11a–11c show

[1] For simplicity, we do not take into account the shift between ω_R and ω_{NBC} due to the repulsion of the frequencies of coupled magnon–photon oscillations in the combined resonator–sample system [17].

examples of these signals for several frequencies of monochromatic pumping. This behavior is similar to the radiation from quasiphonons observed in [14]. Recall that the pumping was performed by a periodic sequence of microwave pulses rather than by a single pulse; however, on the screen of an oscilloscope, all pulses merge into a single one (give identical traces) because the NBC of nuclear magnons is in the same state, i.e., having the same amplitude and phase, after the termination of all microwave pulses.

Now, let us return to the noise pumping. If we switch on the noise modulation of the field H at a fixed frequency of monochromatic pumping, we can simultaneously observe the trailing edges of several pulses on the screen of an oscilloscope; these pulses demonstrate various shapes of radiation signals, similar to those observed under monochromatic pumping for the variation of power. It is obvious that all these traces of the cathode ray of the oscilloscope are attributed to the nonequivalence of the states of the NBC (its amplitude and phase) at the termination moments of different microwave pulses.

Now, let us apply noise pumping with a central frequency of $\omega_p = \omega_R - 2\pi \times 0.2$ MHz and a bandwidth of $\Delta\omega \sim 2\pi \times 1$ MHz. We can simultaneously observe the trailing edges of several pulses on the screen of the oscilloscope (Fig. 11 d). They demonstrate radiation signals of various shapes, similar to those observed under monochromatic pumping for the variation of frequency. This means that, under the noise pumping, not only the amplitude of the NBC (the number of parametric pairs) varies with time, but also the NBC frequency drifts.

The drift of the NBC frequency under noise pumping was also observed in the experiments on investigating the modulation response of the spectrum of a microwave signal transmitted through a resonator. An example of such a spectrum is presented in Fig. 12. The spectrum is recorded with large acquisition time to "average out" fast fluctuations of a microwave signal over frequency. The intense central line in the figure corresponds to a signal of the pumping generator with a spectral bandwidth of $\Delta\omega/2\pi = 40$ kHz, and the side bands represent a modulation response at frequencies $\omega_{NBC} \pm \omega_m$. Since the width of the side bands corresponds to the bandwidth of noise pumping, the range of drift frequencies of the NBC is approximately equal to the pumping bandwidth $\Delta\omega$.

Thus, under noise pumping, an NBC drifts over a wide range of frequencies, while the shape of the electromagnetic signal depends on the frequency of the NBC at the moment when the microwave noise pulse is switched off. However, the amplitude of the maximal signal of radiation (it is shown by curve *2* in Fig. 10) and its relaxation time are virtually the same for the noise and the coherent pumping. This result allows us to conclude that the frequency bandwidth and the amplitude of the NBC of nuclear magnons that is

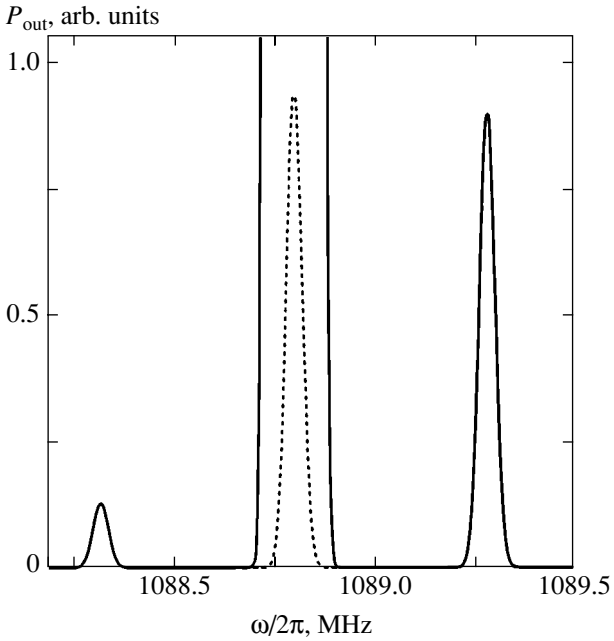


Fig. 12. Spectrum of a microwave signal transmitted through the resonator under noise pumping and coherent modulation of the magnetic field; T = 2.05 K, $\omega_p/2\pi$ = 1088.8 MHz, $\Delta\omega/2\pi$ = 40 kHz, H = 1830 Oe, $\omega_m/2\pi$ = 480 kHz, and P/P_c = 2000. The bandwidth of the spectrum analyzer is 3 kHz, and the time constant is 3 s. The dashed curve represents the central line on the scale 1 : 10000.

formed under noise pumping are virtually the same as those in the case of monochromatic pumping, and that the difference between radiation signals induced by different microwave pulses is mainly associated with the drift of the NBC frequency.

Note one more feature of the electromagnetic radiation of nuclear magnons. This radiation is not observed immediately above the pumping threshold ($P > P_c$) as it was in the system of quasiphonons in $FeBO_3$ [14]. For a sufficiently high supercriticality ($P \sim 1000P_c$), another type of instability develops in a system of NSWs. The mechanism of this instability is unknown. What is important, it is accompanied by a sharp (more than an order of magnitude) increase in the microwave power absorbed by a sample, and hence, by an increase in the NBC amplitude. Electromagnetic radiation of NSWs can be observed precisely above the threshold of this instability. To understand this phenomenon, we write an expression for the radiation power due to nonlinear radiation damping:

$$P_{rad} = \frac{dN_k}{dt} = -\frac{4\pi\hbar Q V^2 N_k^2}{\nu_R}. \qquad (14)$$

Since the radiation power (just as the magnitude of radiation damping) is proportional to the square of the magnetic moment of parametric waves, $P_{rad} \propto V^2 \propto \mu^2$,

and the magnetic moment of NSWs is very small, the radiation of NSWs is very small immediately above the pumping threshold. This radiation becomes comparable to the radiation of magnetoelastic waves observed in [14] at much higher values of supercriticality, when the number of parametric pairs in the NBC becomes greater than the corresponding number N_k^2 of quasiphonons by a factor of greater than 100. Above the new instability threshold of NSWs, the number N_k sharply increases and a radiation signal with a power on the order of the power absorbed by the sample becomes visible. Note that such radiation was also observed under monochromatic pumping, under noise pumping, and under the noise modulation of the spectrum of NSWs.

Thus, we have experimentally demonstrated the formation of a coherent state (nonequilibrium Bose condensate) of pairs of nuclear spin waves under their parametric excitation by a microwave noise field. The existing theory satisfactorily describes the pumping threshold except for the region of low temperatures. The observed discrepancy between theory and experiment can be explained under the assumption about the specific role played by the elastic processes of magnon relaxation. The effect of coherent electromagnetic radiation by a system of parametric pairs of NSWs generated both by monochromatic and noise pumping has been observed for the first time.

ACKNOWLEDGMENTS

I am grateful to V.L. Safonov for discussing the results of the work.

REFERENCES

1. P. G. De Gennes, P. A. Pincus, F. Hartmann-Bourton, and J. M. Winter, Phys. Rev. **129**, 105 (1963).
2. A. V. Andrienko, V. I. Ozhogin, V. L. Safonov, and A. Yu. Yakubovskiĭ, Usp. Fiz. Nauk **161** (10), 1 (1991) [Sov. Phys. Usp. **34**, 843 (1991)].
3. V. S. L'vov, *Nonlinear Spin Waves* (Nauka, Moscow, 1987) [in Russian].
4. A. V. Andrienko, V. L. Safonov, and A. Yu. Yakubovskiĭ, Zh. Éksp. Teor. Fiz. **93**, 907 (1987) [Sov. Phys. JETP **66**, 511 (1987)].
5. A. S. Mikhaĭlov and I. B. Uporov, Zh. Éksp. Teor. Fiz. **77**, 2383 (1979) [Sov. Phys. JETP **50**, 1149 (1979)].
6. V. B. Cherepanov, Fiz. Tverd. Tela (Leningrad) **22**, 43 (1980) [Sov. Phys. Solid State **22**, 25 (1980)].
7. A. V. Andrienko and V. L. Safonov, Pis'ma Zh. Éksp. Teor. Fiz. **60**, 446 (1994) [JETP Lett. **60**, 464 (1994)].
8. A. V. Andrienko, Pis'ma Zh. Éksp. Teor. Fiz. **75**, 79 (2002) [JETP Lett. **75**, 71 (2002)].
9. V. V. Zautkin, B. I. Orel, and V. R. Cherepanov, Zh. Éksp. Teor. Fiz. **85**, 708 (1983) [Sov. Phys. JETP **58**, 414 (1983)].
10. V. L. Safonov, Fiz. Tverd. Tela (St. Petersburg) **34**, 304 (1992) [Sov. Phys. Solid State **34**, 161 (1992)].
11. V. E. Zakharov and V. S. L'vov, Fiz. Tverd. Tela (Leningrad) **14**, 2913 (1972) [Sov. Phys. Solid State **14**, 2513 (1972)].
12. S. A. Govorkov and V. A. Tulin, Zh. Éksp. Teor. Fiz. **91**, 2332 (1986) [Sov. Phys. JETP **64**, 1384 (1986)].
13. V. V. Kveder, B. Ya. Kotyuzhanskiĭ, and L. A. Prozorova, Zh. Éksp. Teor. Fiz. **63**, 2205 (1972) [Sov. Phys. JETP **36**, 1165 (1973)].
14. A. V. Andrienko and V. L. Safonov, Pis'ma Zh. Éksp. Teor. Fiz. **60**, 787 (1994) [JETP Lett. **60**, 800 (1994)].
15. V. L. Safonov, J. Magn. Magn. Mater. **97**, L1 (1991).
16. V. S. Zhitnyuk and G. A. Melkov, Sov. Phys. JETP **48**, 884 (1978).
17. A. V. Andrienko, V. L. Safonov, and H. Yamazaki, J. Phys. Soc. Jpn. **67**, 2893 (1998).

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 100, No. 1, 2005, pp. 89–94.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 1, 2005, pp. 100–106.

ELECTRONIC PROPERTIES OF SOLIDS

Mechanism of Low-Voltage Field Emission from Nanocarbon Materials

Al. A. Zakhidov, A. N. Obraztsov, A. P. Volkov, and D. A. Lyashenko

Moscow State University, Moscow, 119992 Russia
e-mail: obraz@polly.phys.msu.ru

Received April 21, 2004

Abstract—Field emission from nanostructured carbon materials is analyzed by applying the model of emission center in which the emitting surface contains two phases of carbon having substantially different electronic properties. In accordance with this model, the proposed mechanism involves electron tunneling through two potential barriers. The calculated probability of tunneling through two potential barriers implies that the low-voltage field emission observed experimentally can be attributed to the existence of resonant surface states. Numerical estimates suggest that the emission current can increase by at least four orders of magnitude owing to resonant tunneling through two potential barriers. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Field emission (FE) is an important area of basic and applied research [1–3]. Currently, special attention is given to studies of FE from various carbon materials, which are characterized by anomalously low voltages required to generate substantial emission currents. The corresponding threshold strength of the electric field created at the emitting carbon surface is much lower than predicted by the Fowler–Nordheim theory, which adequately describes FE from metals and semiconductors [1, 2]. A comparative analysis of FE from diamond films, carbon nanotubes, and nanostructured carbon materials suggests that the mechanism of the low-voltage emission observed experimentally is universal for all carbonaceous materials and composites and is explained by the possibility of formation of clusters with diamond-like electronic properties on graphite-like carbon surfaces [4]. Apart from data concerning emission characteristics, indirect evidence supporting both this mechanism and the model of emission center is provided by observations of cathode luminescence [5] and photoelectric emission [6].

This mechanism can also be invoked to find consistent explanations for other phenomena involving carbonaceous materials exhibiting low-voltage electron emission, including carbon nanotubes [7–10]; nanocrystalline diamond powders, films, and composites [11–14]; and amorphous diamond-like and related films [15]; and carbon fibers [16]. It should be emphasized that all of these carbonaceous materials exhibit both low-voltage emission and normal FE (as from metals and semiconductors). To generate substantial currents emitted from metals or semiconductors, higher voltages are generally required even in those cases when the geometric characteristics of emitting surfaces (e.g., determined by means of electron microscopy) are analogous to nanocarbon structures (see [17, 18]). It was shown in our recent study [19] that the density of low-voltage emission centers on a carbon cathode surface can be substantially higher (by several orders of magnitude) than the density of electron-emitting centers when a higher voltage is applied. This possibility (attributable to electrostatic screening) impedes interpretation of experimental observations.

In this paper, we present the results of further analysis focused on details of low-voltage emission and on quantitative estimation using the previously proposed phenomenological model of emission center and the corresponding mechanism of low-voltage emission from carbonaceous materials [4–6].

2. PHENOMENOLOGICAL MODEL OF LOW-VOLTAGE EMISSION

The phenomenological model of low-voltage emission center relies on the assumption that the electrons are emitted into a vacuum by tunneling through two potential barriers rather than one (as in the conventional FE theory). It is obvious that the tunneling probability corresponding to a certain combination of barrier parameters and the same applied voltage can be higher in the case of two barriers, as compared to a single one. Analogous twin potential barriers have been analyzed in FE models for cathodes with molecules adsorbed on their surfaces [1, 2] and for cathodes made of thin layers of different materials [2, 20]. Carbon cathodes are special in that twin potential barriers on their surfaces are due to the presence of two different forms of carbon, which have graphite- and diamond-like properties [4–6].

Figure 1 schematizes the energy diagram at the interface between such a nanocarbon cathode and vacuum. In contrast to common representations of dia-

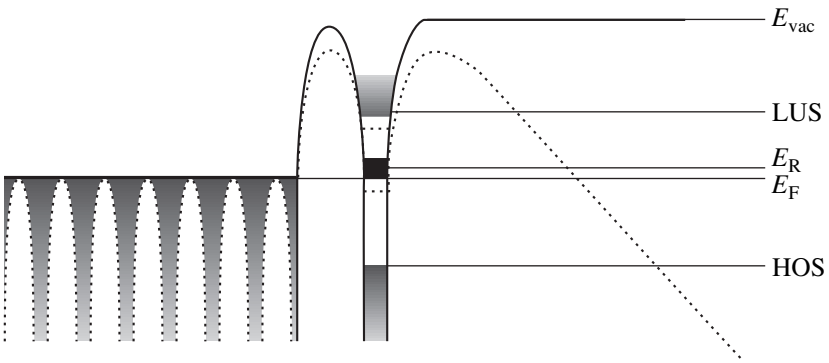


Fig. 1. Schematic energy diagram at the interface between a nanocarbon material and vacuum. Quantum wells correspond to the locations of individual carbon atoms. Gray areas represent energy bands. The Fermi and vacuum levels (E_F and E_{vac}) are represented by horizontal lines. The surface atomic layer is represented by a separate quantum well with two energy bands (HOS and LUS) separated by a LUS–HOS bandgap. Resonant levels (E_R) are shown in the vicinity of the Fermi level. Dotted curves represent energy levels modified by external electric field.

grams of this kind, Fig. 1 shows the quantum wells corresponding to individual carbon atoms located at the surface. Gray areas represent energy bands. For graphite, which can be classified as a semimetal or semiconductor with zero bandgap, the Fermi level (E_F) corresponds to the upper parts of the wells. The surface atomic layer is different in that the energy bands are separated by a gap having a width of about 4 eV [5, 6], which is comparable to the bandgap in a disordered diamond or diamond-like material [15]. The lowest unoccupied and highest occupied states are labeled in Fig. 1 as LUS and HOS, respectively. It is natural to assume that the Fermi level lies in the middle of the LUS–HOS energy gap. This surface layer can tentatively be called diamond-like. A possible mechanism of formation of such diamond-like monolayers on the surfaces of graphite nanocrystallites was discussed in our recent papers [4–6, 21]. Analogous heterogeneity of electronic properties of carbon nanotube surfaces can be attributed to various structural defects [22, 23] or induced by external electric fields [24, 25]. Carbon clusters having diamond-like dielectric properties can manifest themselves by the presence of electrostatic charges on nanotube surfaces [26]. In view of this fact, the energy diagram shown in Fig. 1 can be applied to various nanostructured graphite-like materials, including carbon nanotubes.

Results obtained by means of optical emission spectroscopy suggest that the barrier height at the interface between a diamond-like cluster and vacuum ($E_F - E_{vac}$) is about 4.5 eV [6]. In our model, we set the barrier height between different phases of carbon equal to 4.5 eV, which is slightly lower than the work function for graphite. The thickness of the intermediate layer between the graphite- and diamond-like phases, which determines the corresponding barrier width, is larger in order of magnitude than the interplanar spacing in graphite. In our estimates, both are set equal to 0.4 nm. To simplify calculations, we assume that the inner potential barrier and the barrier at the cathode–vacuum interface (where the external field is applied) have square and wedge-shaped profiles, respectively. Figure 2 shows the simplified energy diagram for the cathode surface.

3. TUNNELING PROBABILITY THROUGH TWIN POTENTIAL BARRIERS

The emission current is determined by probability of tunneling from the cathode material into vacuum. In the most general form, the corresponding current density is expressed as

$$j = e\int v_z T(E_z, F)dk, \quad (1)$$

where k is the electron wavevector, E_z is the electron energy of an electron in the material, F is the electric field strength at the surface, e is the electron charge, $\hbar$ is Planck's constant,

$$v_z = \frac{1}{\hbar}\frac{\partial E_z}{\partial k_z}$$

is the group velocity of an electron, and $T(E_z, F)$ is the transmission coefficient for the barrier. It is obvious that the value of $T(E_z, F)$ characterizing twin barriers differs from that corresponding to a single barrier. To evaluate the transmission coefficient, we first consider the probability of tunneling through each individual barrier separately.

The barrier at the cathode–vacuum interface can be treated in the WKB approximation, as in the conventional FE theory, and approximated by a wedge-shaped profile. Then, the corresponding transmission coefficient is [27]

$$T_2(E_z, F) = P_2^2 = \exp\left[-\frac{4\sqrt{2m}}{3\hbar eF}(E_0 - E_z)^{3/2}\right], \quad (2)$$

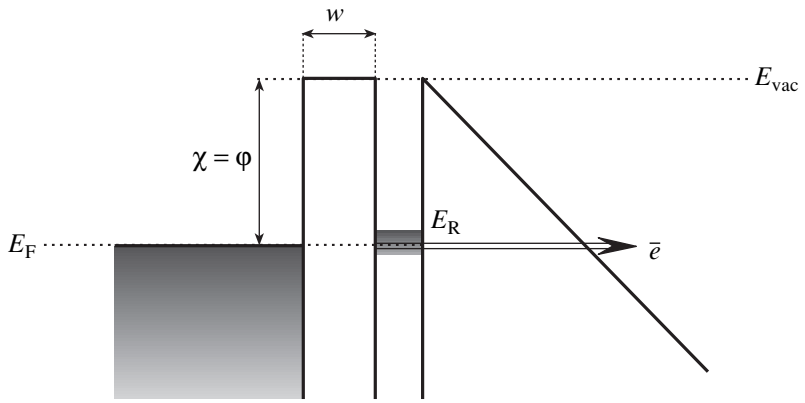


Fig. 2. Simplified schematic representation of the energy diagram at the interface between a nanocarbon material and vacuum with two potential barriers in the presence of external electric field.

where E_0 is the cathode–vacuum barrier height, which is equal to the electron energy in free space relative to the Fermi level. It should be noted here that the actual barrier profile at a matter–vacuum interface is not wedge-shaped, and a rigorous treatment must allow for image forces. However, it was shown in [1, 2] that these forces can be taken into account by introducing a preexponential factor into the final expression for current if the applied field is not very strong.

If the inner potential barrier (at the interface between different carbon phases) is represented by a simple square profile, then the corresponding transmission coefficient is [27]

$$T_1(E_z, F) = P_1^2 = \exp\left[-2w\sqrt{\frac{2m(V-E_z)}{\hbar^2}}\right], \quad (3)$$

where $\chi = V - E_z$ is the effective barrier height and w is the barrier width. The transmission coefficient corresponding to twin square barriers can be expressed as follows (see [28, 29]):

$$T(E_z, F) = \left|P_1P_2\{e^{i\varphi_1} - e^{i\varphi_1'}\} + \frac{P_1}{P_2}\{e^{i\varphi_2} - e^{i\varphi_2'}\} + \frac{P_2}{P_1}\{e^{i\varphi_3} - e^{i\varphi_3'}\} + \frac{1}{P_1P_2}\{e^{i\varphi_4} - e^{i\varphi_4'}\}\right|^{-2}, \quad (4)$$

where P_1 and P_2 denote the respective amplitude transmission coefficients and the expressions in braces are combinations of phase factors. The exact expressions for the phase factors are determined by the barrier heights and widths and the electron wavevector. A general expression analogous to (4) can be used as an approximation in an analysis of tunneling through two barriers one of which is wedge-shaped and the other has a square profile. Since the inner barrier is substantially narrower than the outer one, the transmission coefficient for the former is much greater than for the latter. Accordingly, the sum on the right-hand side of (4) is dominated by the second and fourth terms. In view of the inverse proportionality in (4), this implies that the maximum of $T(E_z, F)$ is reached when the expression in braces in the fourth term vanishes. The conditions corresponding to the highest tunneling probability should be interpreted as resonant tunneling conditions. These conditions are implemented when the energy of a tunneling electron equals a steady-state energy level for an electron in an isolated potential well with parameters (height and width) similar to those of the quantum well between the two barriers considered here [28, 29]. In the case of FE from the surface of a carbonaceous material, resonance conditions are satisfied if the electron energies close to the Fermi level are equal to those of electron states associated with atoms located at the cathode surface.

Surface states are common to various materials. Their parameters, including energy, depend on the material, the surface treatment method, etc. Leaving these factors outside the scope of this study, we note that some surface states can contribute to electron tunneling. It is obvious that the voltage drop associated with a strong external field must lower the surface-state energy levels. This biasing effect can shift the system into or out of a resonant state. In Fig. 1, the resonant states in zero and nonzero external fields are depicted by solid and dotted lines, respectively.

It should be noted that the calculation of the phase factors in braces in expression (4) is a very difficult task requiring additional analysis, including a more accurate determination of the barrier profiles. However, no knowledge of this kind is required in the case of resonant tunneling through two barriers. Previously, analogous resonant mechanisms were discussed with regard to the states that arise on cathode surfaces as a result of adsorption of various molecules [1, 2] and to carbon cathodes (e.g., see [11, 12]). According to the model of

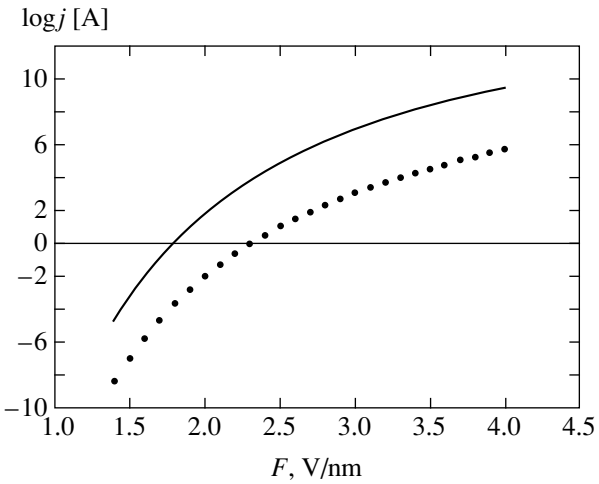


Fig. 3. Tunnel current densities predicted by the Fowler–Nordheim theory (dots) and by expression (5) for low-voltage emission through two potential barriers (solid curve).

emission center discussed in this paper, surface states can be associated with clusters having diamond-like electronic properties that form on the surfaces of well-ordered graphite-like nanoscale structures, such as carbon nanotubes or graphite nanocrystallites. In this respect, the surface states described above are similar in all low-voltage carbon cathodes. Therefore, conclusions based on the present analysis apply to a low-voltage cathode of any type that contains these surface states.

In what follows, we focus on resonant tunneling with $\varphi_4 - \varphi_4' = 2\pi n$, where n is an integer. In this case, the fourth term (classical analog of the tunneling probability through two barriers given by the product of the respective tunneling probabilities) can be neglected, and only the second term is retained. It can be shown that the phase differences in the second and fourth terms in (4) cannot vanish simultaneously [28, 29]. Then, taking into account image forces, we can use expression (4) to derive the emission current density

$$j = \frac{e^3}{8\pi h}\frac{1}{\left(\frac{\sqrt{\varphi}}{eF} - \frac{w}{2\sqrt{\chi}}\right)^2} \times \exp\left(1.03\frac{8\pi e^2}{3h}\sqrt{\frac{2m}{\varphi}}\right) \times \exp\left(-0.95\frac{4\pi\sqrt{2m}}{h}\left(\frac{2\varphi^{3/2}}{3eF} - \sqrt{\chi}w\right)\right). \quad (5)$$

This expression differs from that obtained in the conventional FE theory (see [1–3]) by additional summands in the preexponential factor and exponent. Using the parameters adopted above for our phenomenological model ($\chi = 4.5$ eV and $w = 4$ Å), we find that the following relation holds for $F < 5 \times 10^9$ V/m:

$$\frac{w}{2\sqrt{\chi}}\Big/\frac{\sqrt{\varphi}}{eF} \ll 1.$$

Thus, if the inner barrier is narrower than the outer one, then this quantity contained in the preexponential factor does not significantly modify the expression for current density as compared to the Fowler–Nordheim formula.

However, the value given by (5) strongly depends on the summand added to $(\chi^{1/2}w)$ in the exponent. Substituting the parameters specified above into (5), we see that the emission current density can exceed those predicted by the Fowler–Nordheim theory for emission from metals or semiconductors. Figure 3 compares the field-dependent current densities calculated by using expression (5) and the Fowler–Nordheim formula.

These results are obtained for an isolated emission center, whereas electrons are actually emitted by numerous emission centers located on the cathode surface. Since the geometry of emission centers varies, the mean electric field strengths corresponding to the emission threshold vary accordingly [19]. The number of emission centers on a metal or semiconductor cathode surface must increase with voltage as local electric field strengths progressively exceed the threshold values determined by local values of work function. The highest emission current density is reached when the mean field strength is much higher than the local threshold values, depending on the free-carrier concentration and mobility, whereas the density of centers where emission is controlled by the mechanism discussed here may decrease with increasing field strength. The latter trend is explained by the possible saturation of the emission current due to a limited density of resonant surface states (whose number is obviously less than the number of free electron in the cathode material). For example, this effect is characteristic of emission from nanostructured-carbon film cathodes [19]. This behavior of carbonaceous cathodes at low voltages leads to a total emission current from macroscopic cathode surfaces containing numerous emission centers that is even higher than that predicted by using expression (5) for current density.

To elucidate the role played by the inner barrier, let us consider the ratio of the resonant emission current I_R to the conventional field-emission current I_{FN} given by the Fowler–Nordheim formula as a function of the inner barrier width w:

$$\frac{I_R}{I_{FN}} = \frac{\varphi}{\left(\sqrt{\varphi} - \frac{eFw}{2\sqrt{\chi}}\right)^2} \times \exp\left(0.95\frac{4\pi\sqrt{2m}}{h}\sqrt{\chi}w\right). \quad (6)$$

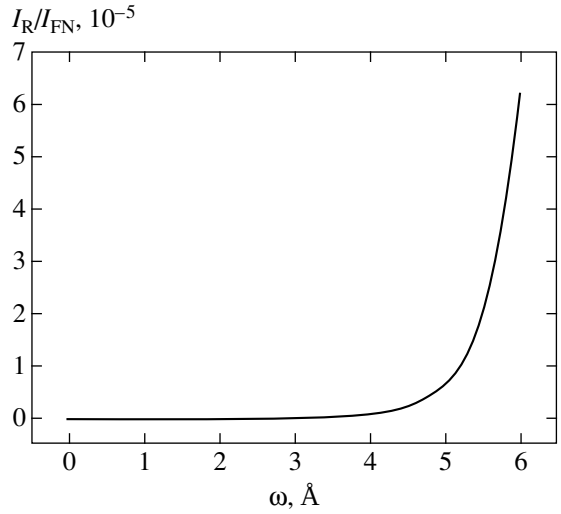


Fig. 4. Ratio of resonant field-emission current to the emission current given by the Fowler–Nordheim formula versus inner barrier width w for $F = 2$ V/nm applied to an emission center.

Figure 4 shows this function calculated for $F = 2 \times 10^9$ V/m. It should be noted that estimates of this kind are valid if w is much narrower than the outer barrier width, i.e., if the inner barrier is easily passed by electrons. The dependence of the emission current on the barrier width illustrates the fact (paradoxical from a classical-mechanics standpoint) that an increase in the width of one potential barrier results in an exponential increase in the tunneling probability through two barriers. This quantum-mechanical effect is due to the existence of at least one quasi-steady resonant level between the barriers. As noted above, the number of resonant states contributing to the emission mechanism considered here depends on the presence of certain clusters having modified electronic properties on a carbon cathode surface. When the density of these resonant states is much lower than the density of the free carriers that may contribute to the emission current, the total emission current is strictly limited even though the tunneling probability may increase exponentially. A more detailed analysis of upper limits on resonant tunneling current should be based on further experimental and theoretical studies of resonant states on the surfaces of nanocarbon materials.

4. CONCLUSIONS

According to the analysis presented above, the existence of resonant surface states can explain the mechanism of low-voltage field emission from nanocarbon materials. One essential requirement is the heterogeneity of the surface properties of a curved graphite-like atomic layer, which gives rise to twin potential barriers. Accordingly, low-voltage emission of electrons from various nanocarbon materials is characterized by a sharply resonant behavior and can be implemented when surface states occupy certain relative positions. The current density due to low-voltage emission is an exponential function of the thickness of the surface layer having modified electronic properties. Effective resonant tunneling requires that this thickness be comparable to the width of the potential barrier at the material–vacuum interface. Formation of such narrow atomic layers is a unique characteristic of carbonaceous materials due to the fact that carbon atoms can organize into well-conducting graphite-like and dielectric diamond-like structures.

ACKNOWLEDGMENTS

This work was supported by INTAS, project no. 01-0254 and CRDF, project no. RP2-2559-MO-03.

REFERENCES

1. *Nonincandesced Cathodes*, Ed. by M. I. Elinson (Sovetskoe Radio, Moscow, 1974) [in Russian].
2. R. Gomer, *Field Emission and Field Ionization* (AIP, New York, 1993).
3. D. Temple, Mater. Sci. Eng. R **24**, 185 (1999).
4. A. N. Obraztsov, A. P. Volkov, and I. Yu. Pavlovskiĭ, Pis'ma Zh. Éksp. Teor. Fiz. **68**, 56 (1998) [JETP Lett. **68**, 59 (1998)].
5. A. P. Volkov, A. N. Obraztsov, I. Yu. Pavlovskiĭ, *et al.*, Poverkhnost, Nos. 5–6, 161 (1999).
6. A. N. Obraztsov, A. P. Volkov, A. I. Boronin, and S. V. Koshcheev, Zh. Éksp. Teor. Fiz. **120**, 970 (2001) [JETP **93**, 846 (2001)].
7. J.-M. Bonard, Th. Stoeckli, F. Maier, *et al.*, Phys. Rev. Lett. **81**, 1441 (1998).
8. K. A. Dean, P. von Allmen, and B. R. Chalamala, J. Vac. Sci. Technol. B **14**, 1959 (1999).
9. Z. L. Wang, R. P. Gao, W. A. de Heer, and P. Poncharal, Appl. Phys. Lett. **80**, 856 (2002).
10. Y. Chen, D. T. Shaw, and L. Guo, Appl. Phys. Lett. **76**, 2469 (2000).
11. A. V. Karabutov, V. G. Ralchenko, I. I. Vlasov, *et al.*, J. Vac. Sci. Technol. B **21**, 597 (2003).
12. A. V. Karabutov, V. D. Frolov, and V. I. Konov, Diamond Relat. Mater. **10**, 840 (2001).
13. J. W. Steeds, A. Gilmore, K. M. Bussmann, *et al.*, Diamond Relat. Mater. **8**, 996 (1999).
14. J. B. Cui, M. Stammler, J. Ristein, and L. Ley, J. Appl. Phys. **88**, 3667 (2000).
15. J. Robertson, Thin Solid Films **296**, 61 (1997).
16. R. V. Latham and D. A. Wilson, J. Phys. D: Appl. Phys. **14**, 2139 (1981).
17. K. W. Wong, X. T. Zhou, F. C. K. Au, *et al.*, Appl. Phys. Lett. **75**, 2918 (1999).
18. M. Kanechika and Y. Mitsushima, Jpn. J. Appl. Phys. **39**, 7111 (2000).

19. Al. A. Zakhidov, A. N. Obraztsov, A. P. Volkov, and D. A. Lyashenko, Zh. Éksp. Teor. Fiz. **124**, 1391 (2003) [JETP **97**, 1240 (2003)].

20. V. T. Binh and Ch. Adessi, Phys. Rev. Lett. **85**, 864 (2000).

21. A. N. Obraztsov, A. P. Volkov, I. Yu. Pavlovskiĭ, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 381 (1999) [JETP Lett. **69**, 411 (1999)].

22. K. Suenaga, C. Colliex, and S. Iijima, Appl. Phys. Lett. **78**, 70 (2001).

23. Z. Klusek, P. Kowalczyk, and P. Byszewski, Vacuum **63**, 145 (2001).

24. C. Kim, B. Kim, S. M. Lee, *et al.*, Appl. Phys. Lett. **79**, 1187 (2001).

25. T. Kuzumaki, Y. Takamura, H. Ichinose, and Y. Horiike, Appl. Phys. Lett. **78**, 3699 (2001).

26. Z. L. Wang, R. P. Gao, W. A. de Heer, and P. Poncharal, Appl. Phys. Lett. **80**, 856 (2002).

27. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory*, 4th ed. (Nauka, Moscow, 1989; Pergamon, New York, 1977).

28. E. O. Kane, J. Appl. Phys. **32**, 83 (1961).

29. E. O. Kane, in *Tunneling Phenomena in Solids*, Ed. by E. Burstein and S. Lundqvist (Plenum, New York, 1969; Mir, Moscow, 1973).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 100, No. 1, 2005, pp. 95–105.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 1, 2005, pp. 107–119.

ELECTRONIC PROPERTIES OF SOLIDS

Peculiarities of the Magnetic State in the System $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ ($0 \le \gamma \le 0.25$)

S. V. Trukhanov

State Research Institute of Solid-State and Semiconductor Physics, National Academy of Sciences of Belarus, Minsk, 220072 Republic of Belarus

e-mail: truhanov@ifttp.bas-net.by

Received July 1, 2004

Abstract—The results of experimental investigation of the chemical phase composition, crystal structure, and magnetic properties of the manganite $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ ($0 \le \gamma \le 0.25$) with perovskite structure depending on the concentration of oxygen vacancies are presented. It is found that the mean grain size of the stoichiometric solid solution of $La_{0.70}Sr_{0.30}MnO_3$ amounts approximately to 10 μm, while the grain size for anion-deficient solid solutions of $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ is approximately 5 μm. It is found that samples with $0 \le \gamma \le 0.13$ have a rhombohedral unit cell (with space group $R\bar{3}c$, $Z = 2$), while samples with $\gamma \ge 0.20$ have a tetragonal unit cell (space group $I4/mcm$, $Z = 2$). It is proved experimentally that the magnetic phase state of the manganite $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ changes upon a decrease in the oxygen content. It is shown that anion-deficient solid solutions of $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ experience a number of successive magnetic phase transformations in the ground state from a ferromagnet ($0 \le \gamma \le 0.05$) to a charge-disordered antiferromagnet ($\gamma = 0.25$) via an inhomogeneous magnetic state similar to a cluster spin glass ($0.13 \le \gamma \le 0.20$). The mean size of ferromagnetic clusters ($r \approx 50$ nm) in the spin glass state is estimated. It is shown that oxygen vacancies make a substantial contribution to the formation of magnetic properties of manganites. The generalized magnetic characteristics are presented in the form of concentration dependences of the spontaneous magnetic moment, coercive force, and the critical temperature of the magnetic transition. The most probable mechanism of formation of the magnetic phase state in Sr-substituted anion-deficient manganites is considered. It is assumed that in the absence of orbital ordering, a decrease in the magnetic ion coordination number leads to sign reversal in indirect superexchange interactions Mn^{3+}–O–Mn^{3+}. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The discovery of giant magnetoresistance (GMR) in orthomanganites with the general chemical formula $Ln_{1-x}A_xMnO_3$ (Ln stands for a trivalent rare-earth cation and A is a bivalent alkali-earth cation) stimulated intense studies of their physicochemical properties [1–4]. The essence of the GMR effect is that the resistance of substituted manganites sharply decreases at the instant of phase transitions to the ferromagnetic state after the application of an external magnetic field. The magnetoresistance of manganites is many orders of magnitude higher than that for multilayered films and granular systems; this effect is observed at room temperatures and in weak magnetic fields [5]. This opened wide prospects for practical applications of manganites. Phase transformations associated with orbital and charge states, which determine the type of magnetic ordering, are also an interesting feature of manganites. For example, charge-ordered compounds are antiferromagnetic insulators as a rule, while charge-disordered compounds are ferromagnetic metals [6].

It is well known that magnetic and electric states of manganites are mainly determined by the concentration ratio of tri- and tetravalelent ions of manganese. This ratio can usually be changed by doping with an alkali-earth metal. However, other methods such as substitution of magnetic and nonmagnetic ions with different valences for manganese or the use of nonstoichiometric oxygen also exist [7]. For example, it was found that oxidation of $LaMnO_3$ leads to a transition from the antiferromagnetic state to the ferromagnetic state in which the interdependence of magnetic and electrical properties is clearly manifested [8]. The emergence of oxygen vacancies leads to a sharp increase in the ion conductivity in manganites, which is interesting for technical applications [9].

Substituted manganites are good model objects for studying physical properties of strongly correlated electronic systems. However, the general level of the theory does not permit an adequate description of all the results of experimental studies. The nature of magnetic and electrical phenomena in manganites remains the subject of heated discussion [10–12].

The compound $LaMnO_3$ forms the basis of all substituted lanthanum manganites. This compound is an antiferromagnetic semiconductor with a magnetic structure of the *A* type [13, 14]. The weak ferromagnetic component is due to antisymmetric Dzialoshin-

sky-Moriya exchange [15, 16]. The Néel temperature for this compound is approximately 140 K. It should be noted that the magnetic properties of all manganites are determined by the spins of manganese ions since the orbital magnetic moment of these ions are "frozen" in the crystal field of oxygen anions. The presence of Jahn–Teller Mn^{3+} ions with electron configuration $t_{2g}^3 e_g^1$ ($S = 2$), which appear in the stoichiometric compound $La^{3+}Mn^{3+}O_3^{2-}$, is responsible for the *O*' orthorhombic ($c/\sqrt{2} < a \le b$) symmetry of the unit cell of this compound. In the octahedral oxygen surroundings, the *d* shell of the Mn^{3+} ion splits into a doublet and a triplet. In the high-spin state, the triply degenerate level t_{2g} is completely filled with three electrons, while the doubly degenerate level e_g contains only one electron. Such a high-spin configuration is associated with a strong Hund exchange coupling, which orients all electron spins in the same direction. It is assumed that e_g^1 electrons are localized at manganese ions and cannot move freely over the crystal lattice.

The dependence of the magnetic properties of substituted manganites with perovskite structure on oxygen stoichiometry has not been studied extensively. As a result of removal of oxygen anions from the crystal lattice of a solid solution $La_{1-x}A_x^{2+}Mn_{1-x+2\gamma}^{3+}Mn_{x-2\gamma}^{4+}O_{3-\gamma}^{2-}$, the Mn^{4+} ions are converted into Mn^{3+} ions and their coordination number decreases from 6 to 5 [17]. Of all the 3*d*-metal-based compounds, manganites are characterized by the widest interval of variation of oxygen stoichiometry. For example, the oxygen index for manganites may vary from 2.5 [18] to 3.29 [19]. Titanates and vanadates a much smaller deviation from stoichiometry (from 3 to 3.08 and 3 to 3.05, respectively) [20]. Oxygen vacancies in manganites can be ordered, as was observed in the case of $CaMnO_{3-\gamma}$; this leads to interesting peculiarities in the crystal structure and magnetic properties, such as an increase in T_C and μ_{eff} [21].

The Sr-substituted manganite $La_{0.70}Sr_{0.30}MnO_3$ was chosen as the object for investigation in this study. We analyzed the influence of oxygen stoichiometry on its crystal structure and magnetic properties. Earlier [22], a number of $La_{1-x}Sr_xMnO_{3-x/2}$ anion-deficient solid solutions, which do not contain Mn^{4+} ions, were investigated. Such a situation was realized by a matched variation of two parameters, viz., the concentration of (1) substituent cations Sr^{2+} and (2) O^{2-} anions. Interpretation of the magnetic properties of $La_{1-x}Sr_xMnO_{3-x/2}$ cannot be based on the double-exchange mechanism model due to the absence of heterovalent Mn ions. In the system $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$, the average valence of manganese ions gradually decreases, leading to the formation of 20% Mn^{2+} ions for $\gamma = 0.25$. Thus, it is interesting to follow the dynamics and, whenever possible, separate the effects of different contributions (double exchange and superexchange) to the formation of the magnetic state of anion-deficient manganites with a constant concentration of Sr^{2+} cations. A comparative analysis of the results of investigation of two systems, $La_{1-x}Sr_xMnO_{3-x/2}$ and $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$, indicates that the concentration of oxygen vacancies is an additional parameter significantly contributing to the formation of magnetic properties of systems with the same chemical composition and varying valences of manganese ions.

Analysis of the crystal structure, magnetic and electrical properties of anion-deficient compound $La_{0.70}Sr_{0.30}MnO_{3-\delta}$ was carried out in [23]. It was found that a transition to orbitally ordered *O*' orthorhombic structure takes place in the concentration range $0.075 \le \delta \le 0.1$; this is quite astonishing since the presence of oxygen vacancies must break the symmetry in the arrangement of the *d* orbitals of manganese ions and hamper the orbital ordering. The samples below T_C are ferromagnets for $0 \le \delta \le 0.1$ and orbitally ordered antiferromagnets for $0.125 \le \delta \le 0.15$. As the oxygen deficiency δ decreases, the Curie temperature also decreases together with spontaneous magnetization. It was also found that the resistivity gradually increases and becomes semiconductor-type, while the temperature of transition to magnetically ordered state is displaced towards low temperatures. However, no mention is made in [23] about the temperature- and magnetic field dependences of magnetization. The magnetic transition temperature was determined in a quite strong magnetic field of 0.5 kOe; the structural data are also inadequate. However, the physical properties of manganites are very sensitive to the method of their preparation and the unit cell symmetry. Hence we present here detailed experimental results on the chemical phase composition, crystal structure, and magnetic properties of Sr-substituted anion-deficient manganite $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ ($0 \le \gamma \le 0.25$).

2. EXPERIMENT

The method for obtaining polycrystalline stoichiometric $La_{0.70}Sr_{0.30}MnO_3$ and anion-deficient $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ solid solutions is given in [22]. X-ray diffraction analysis was carried out on a diffractometer DRON-3 in the K_α radiation of Cr at room temperature in the angular interval $20° \le 2\theta \le 100°$ with a scanning step $\Delta\theta = 0.03$. A graphite monochromator was used to filter out the K_β radiation. The oxygen concentration was determined by thermogravimetric analysis (TGA). The absolute error in determining the oxygen index was ±0.01, and the chemical formula of anion-deficient solid solutions can be written as $La_{0.70}Sr_{0.30}MnO_{3-\gamma\pm0.01}$. A significant feature of anion-deficient perovskite-like solid solutions obtained by the method of topotactic reactions is the possibility of their oxidation accompanied by restoration of the initial composition, structure, and physical properties [24].

The microstructure and chemical phase composition of the samples were analyzed on a NANOLAB-7 scanning electron microscope containing two mutually complementary X-ray analyzers: MS-46 (wave dispersion spectrometer) and System 860-50 (energy dispersion spectrometer). The samples were degreased before the experiments by washing in ethyl alcohol in an ultrasonic bath for approximately 5 min.

The magnetic properties of anion-deficient samples $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ were studied on a commercial OI-3001 vibrating-coil magnetometer in the temperature range 4.2–350 K. The temperature dependences of the specific magnetic moment were measured in weak fields (up to 100 Oe) during heating after zero-field cooling (ZFC) and in the field cooling (FC) mode; the field dependences at low temperatures (5 K) were also measured. The magnetic ordering temperature T_{mo} was determined from the temperature dependence of the FC-curve for specific magnetic moment in a weak magnetic field of 100 Oe as the temperature corresponding to the sharpest decrease in the specific magnetic moment ($\min\{dM_{FC}/dT\}$). For samples with a blurred magnetic phase transition, T_{mo} was determined as the temperature at which the specific magnetic moment became about 10% smaller than its maximal value. The freezing temperature T_f for magnetic moments of ferromagnetic clusters was defined as the temperature corresponding to the peak on the ZFC curve for the specific magnetic moment. Spontaneous atomic magnetic moment M_s was determined from the field dependence by linear extrapolation to zero field.

3. EXPERIMENTAL RFESULTS AND DISCUSSION

The phase composition and the unit cell parameters were determined by X-ray diffraction technique using the anion deficient samples of $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ as objects of investigation. Powder diffraction patterns for some solid solutions are shown in Fig. 1. It was found that all samples had perovskite structure. Within the experimental error (less than 3%), impurity phases were not detected.

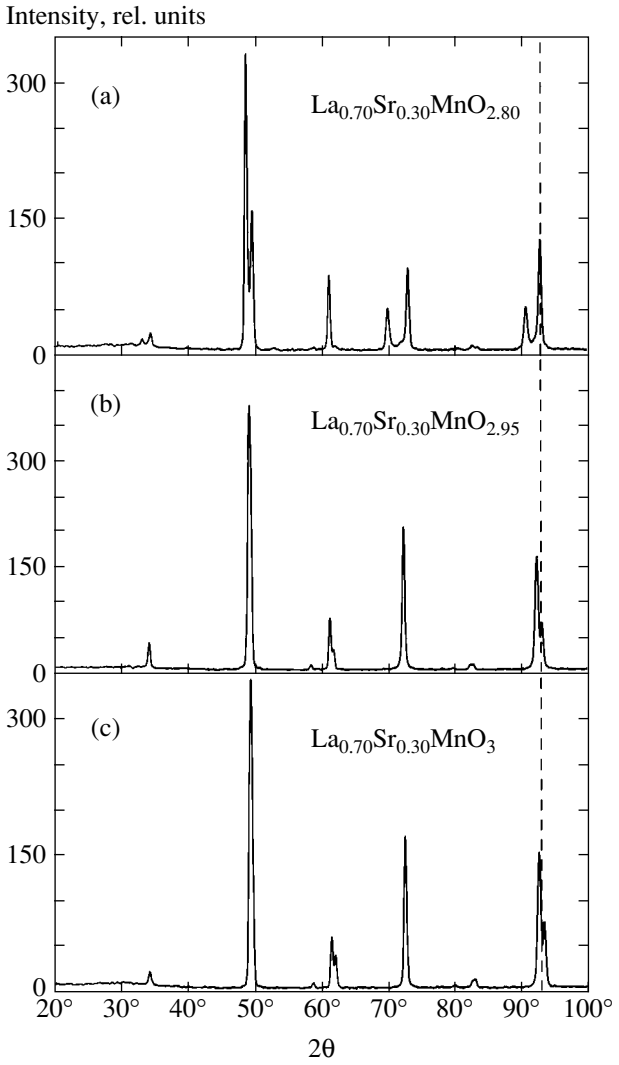


Fig. 1. Powder X-ray diffraction patterns obtained at room temperature for the following solid solutions: (a) anion-deficient $La_{0.70}Sr_{0.30}MnO_{2.80}$; (b) $La_{0.70}Sr_{0.30}MnO_{2.95}$, and (c) stoichiometric $La_{0.70}Sr_{0.30}MnO_3$.

An important condition for the emergence of the ABO_3 perovskite structure is a favorable cation size. One of the cations (A) must always be larger and must have a radius nearly the same as the radius of the oxygen anion, while the other cation (B) in the octahedral coordination must be substantially smaller. The radii of cations A and B must be within the limits roughly set, on the one hand, by the conditions according to which the coordination number must be 6 or 12. On the other hand, these limits are determined by the so-called stability (tolerance) factor t defined by Goldsmith [25] as follows:

$$t = \frac{\langle r(\mathrm{A})\rangle + \langle r(\mathrm{O})\rangle}{\sqrt{2}(\langle r(\mathrm{B})\rangle + \langle r(\mathrm{O})\rangle)}; \quad (1)$$

here, $\langle r(\mathrm{A})\rangle$, $\langle r(\mathrm{B})\rangle$, and $\langle r(\mathrm{O})\rangle$ are, respectively, the mean radii of the ions in the A and B positions in perovskite ABO_3 and of the oxygen ion.

Oxide compounds have perovskite structure if their tolerance factor lies in the limits of $0.82 < t < 1.02$. The value of t must be equal to unity in the ideal case. However, systematic study of oxides with perovskite structure revealed that the value of t is slightly lower in actual practice. The limits of the numerical values of this parameter (which determine the range of the perovskite structure) are determined by the valence of cations and are found, for example, to be $t_{min} = 0.77$ and $t_{max} = 0.99$ for a perovskite of the $CaMnO_3$ type. If the radius of ion A is so small that $t < t_{min}$, a corundum or ilmenite structure is formed most frequently. For $t > t_{max}$, a calcite or aragonite structure is formed [26]. The

Goldsmith's tolerance factor (t), oxygen vacancy concentration (γ), concentration of trivalent manganese ions, symmetry type (S), and unit cell parameters (a, c, α, and V) for anion-deficient solid solutions $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ ($0 < \gamma \le 0.25$)

Chemical composition	t, Å	γ	Mn^{3+}, %	R	a, Å	c, Å	α	V, Å^3
$La_{0.70}Sr_{0.30}MnO_3$	1.0021	0	70	R	5.464		60.44°	116.82
$La_{0.70}Sr_{0.30}MnO_{2.95}$	0.9766	0.05	80	R	5.478		60.31°	117.39
$La_{0.70}Sr_{0.30}MnO_{2.87}$	0.9727	0.13	96	R	5.501		60.09°	117.92
$La_{0.70}Sr_{0.30}MnO_{2.80}$	0.9675	0.20	90	T	3.959	7.642		119.79
$La_{0.70}Sr_{0.30}MnO_{2.75}$	0.9649	0.25	80	T	3.978	7.644		120.98

Note: R stands for the rhombohedral symmetry with space group $R\bar{3}c$ and T indicates the tetragonal symmetry with space group $I4/mcm$. The values of the effective ionic radii are borrowed from [28]: $r[La^{3+}(12)] = 1.360$ Å, $r[La^{3+}(10)] = 1.270$ Å, $r[Sr^{2+}(12)] = 1.440$ Å, $r[Sr^{2+}(10)] = 1.360$ Å, $r[Mn^{2+}(5)] = 0.750$ Å, $r[Mn^{3+}(6)] = 0.645$ Å, $r[Mn^{3+}(5)] = 0.580$ Å, $r[Mn^{4+}(6)] = 0.530$ Å.

values of t for anion-deficient solid solutions $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ studied here are given in the table.

An ideal perovskite unit cell possesses cubic symmetry with space group $Pm3m$. However, compounds with cubic perovskite structure are encountered quite rarely. The crystal lattice of a perovskite is usually distorted as a result of (i) mismatching between the sizes of the cations and pores occupied by them and (ii) the Jahn–Teller effect (JTE). In the former case, a transition to the free-energy minimum is attained due to cooperative rotation of oxygen octahedrons MnO_6. Tetrahedral distortions ($a = b \neq c$, $\alpha = \beta = \gamma$) appear as a result of rotation of the initial cubic lattice about the [100] axis, orthorhombic distortions ($a \neq b \neq c$, $\alpha = \beta = \gamma$) appear as a result of rotation about the [110] axis, and rhombohedral distortions take place when the structure rotates about the [111] axis ($a = b = c$, $\alpha = \beta = \gamma \neq 90°$). In the second case, distortions appear as a result of removal of degeneracy of e_g electron levels of the Jahn-Teller Mn^{3+} ions in the octahedral field of oxygen anions [24].

The unit cell symmetry in anion-deficient solid solutions $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ changes with the oxygen vacancy concentration. Samples with $0 \le \gamma \le 0.13$ possess a rhombohedral unit cell (space group $R\bar{3}c$, $Z = 2$), while samples with $\gamma \ge 0.20$ have a tetragonal unit cell (space group $I4/mcm$, $Z = 2$). It is worth noting that such X-ray reflections are displaced towards smaller Bragg angles (Fig. 1) upon an increase in the oxygen deficiency, which corresponds to an increase in the unit cell volume.

It was found in [23] that the anion-deficient solid solution $La_{0.70}Sr_{0.30}MnO_{3-\delta}$ in the concentration range $0.075 \le \delta \le 0.1$ experiences a transition to the orbitally ordered O' orthorhombic structure. This conclusion is dubious since the presence of oxygen vacancies must break the symmetry in the arrangement of d orbitals of the nearest manganese ions and hamper the orbital ordering. According to Goodenough [27], the O' orthorhombic symmetry ($c/\sqrt{2} < a \le b$) is associated with the cooperative static JTE, resulting in antiferrodistortion ordering of d_{z^2} orbitals of Mn^{3+} ions. The JTE means a decrease in the free energy of a degenerate electron system due to a decrease in its symmetry as a result of removal of electron level degeneracy. The necessary condition for the emergence of the JTE is the presence of a degenerate electron system. These can be d^9 and d^7 ions in a low-spin state or d^4 ions in a high-spin state. In the case of manganites, the degeneracy of the e_g levels of Mn^{3+} in the octahedral surrounding of oxygen anions is removed. For the $La_{1-x}Ca_xMnO_{3-\gamma}$ system with $x \sim 0.1$, a transition from O' orthorhombic to O orthorhombic symmetry ($a < c/\sqrt{2} < b$) was detected earlier [28]. The unit cell parameters of anion-deficient solid solutions $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ considered here are given in the table. The presence of the tetragonal symmetry in samples with $\gamma \ge 0.20$ indicates the absence of orbital ordering in them.

Figure 2 shows the results of investigations by scanning electron microscopy. These results show that a stoichiometric sample of $La_{0.70}Sr_{0.30}MnO_3$ is characterized by a mean grain size of about 10 μm, while anion-deficient sample of $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ have a grain size of about 5 μm. In this study, the grains are homogeneous spatial regions separated by continuous boundaries. It can be seen from the figure that the grain size is characterized by a considerable spread. Thus it can be concluded that the thermal treatment of polycrystalline samples at lower temperatures decreases the mean size of the grains and increases their porosity. It is also found that all the samples exhibit uniform distribution of chemical elements corresponding to the basic chemical formula $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$.

As the oxygen deficiency γ increases, the unit cell volume for anion-deficient samples $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ increases monotonically (see table). Such a behavior is due to the fact that the concentration of Mn^{3+} ions

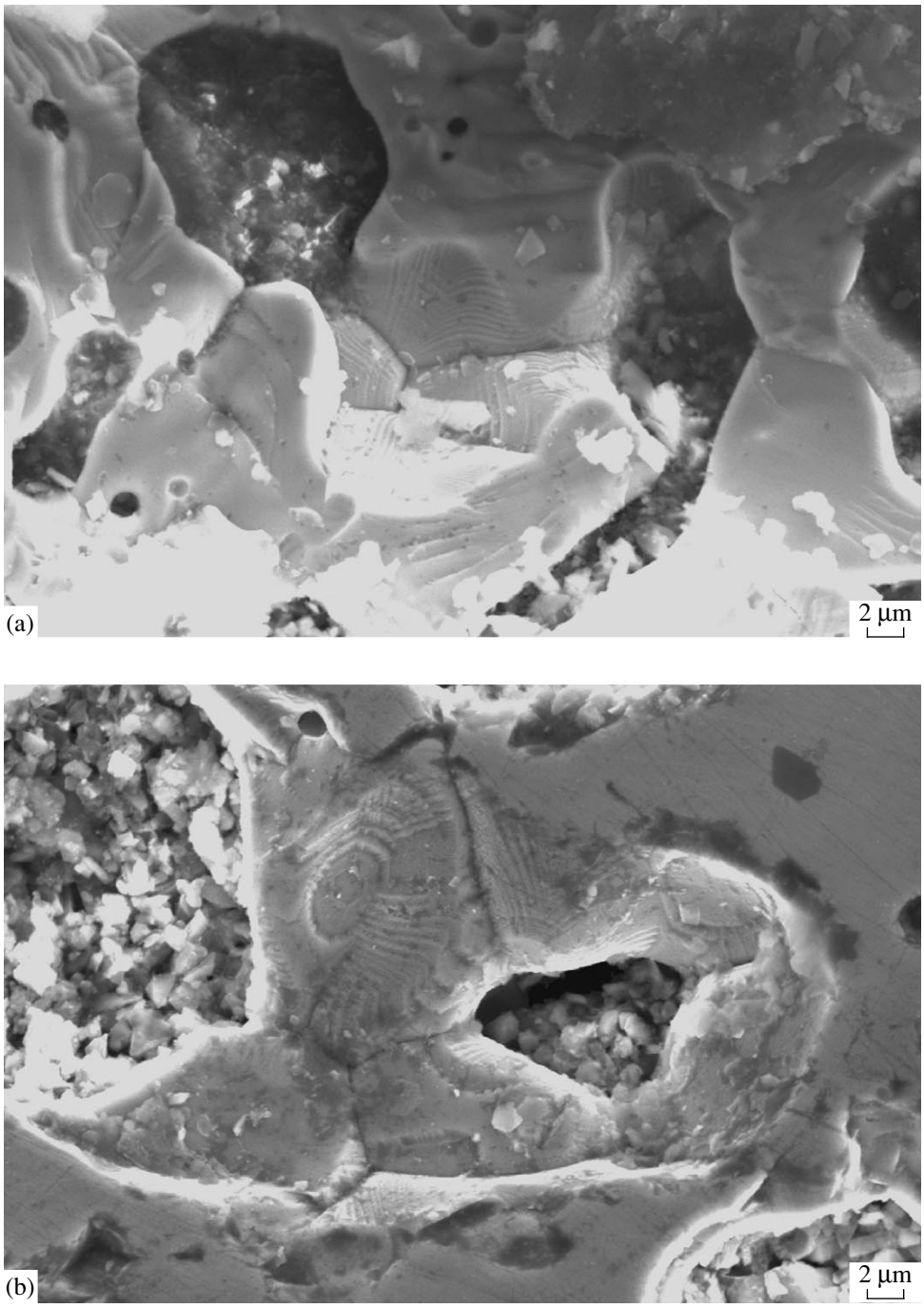


Fig. 2. Photographs obtained using a scanning electron microscope for solid solutions of (a) stoichiometric compound $La_{0.70}Sr_{0.30}MnO_3$ and (b) anion-deficient compound $La_{0.70}Sr_{0.30}MnO_{2.80}$.

increases in the system ($r\{Mn^{3+}(6)\}$ = 0.645 Å). The effective radius of these ions in the octahedral surrounding of oxygen anions is larger than that of Mn^{4+} ions ($r\{Mn^{4+}(6)\}$ = 0.530 Å). The emergence of oxygen vacancies leads to a decrease in the coordination number and hence to a decrease in the effective radius of Mn^{3+} ions. Oxygen vacancies are formed precisely near Mn^{3+} ions since the force of electrostatic repulsion between these ions is smaller than in the case of Mn^{4+} ions. The effective radius of $Mn^{3+}(5)$ in the pentahedral coordination is 0.580 Å [29].

Figure 3 shows the temperature dependence of the specific magnetic moment for some anion-deficient solid solutions of $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ in the ZFC and FC modes. The stoichiometric compound $La_{0.70}Sr_{0.30}MnO_3$ is a ferromagnet [22] with a Curie

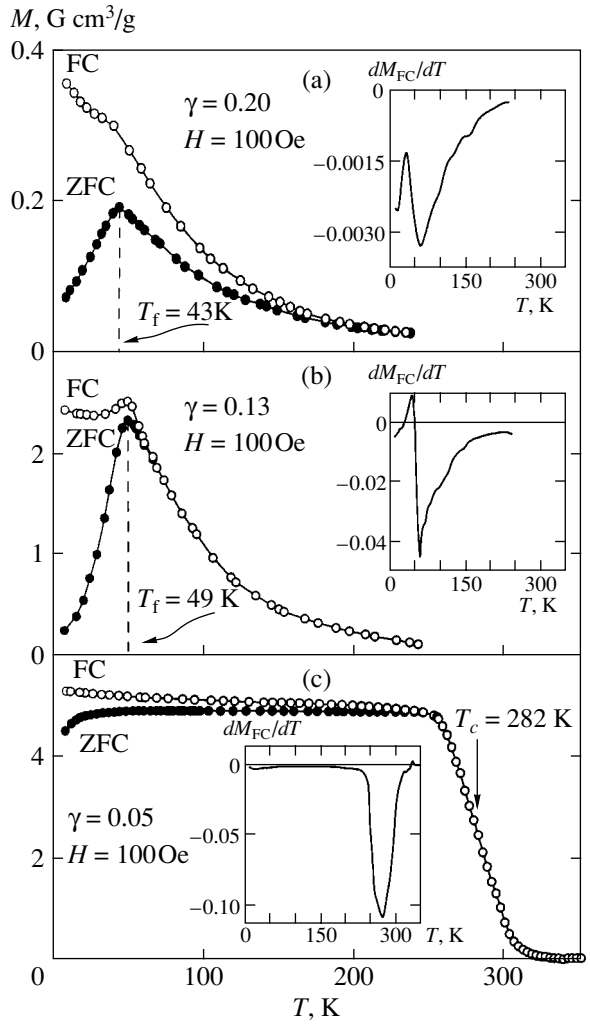


Fig. 3. Temperature dependence of the specific magnetic moment in a field of 100 Oe, measured upon increase in temperature after zero-field cooling (ZFC) (dark symbols) and in a field of 100 Oe (FC) (light symbols) for anion-deficient solid solutions of (a) $La_{0.70}Sr_{0.30}MnO_{2.80}$, (b) $La_{0.70}Sr_{0.30}MnO_{2.87}$, and (c) $La_{0.70}Sr_{0.30}MnO_{2.95}$. The inset demonstrates the temperature dependence of the derivative of the FC curve.

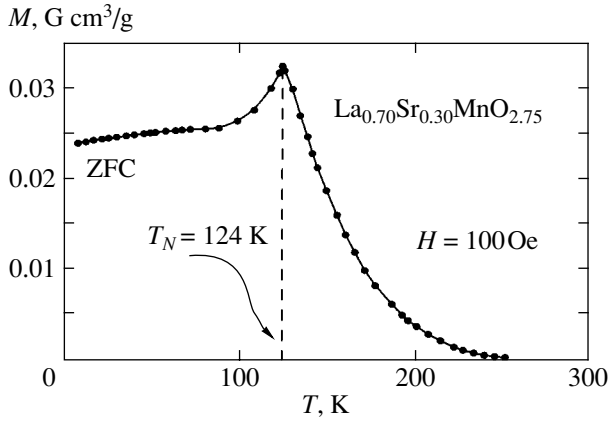


Fig. 4. Temperature dependence of the ZFC curve describing the specific magnetic moment in a field of 100 Oe for the anion deficient solid solution $La_{0.70}Sr_{0.30}MnO_{2.75}$.

temperature of about 350 K and a spontaneous magnetic moment on the order of $3.66\mu_B$ per structural unit. The Curie temperature gradually decreases with increasing concentration of oxygen vacancies. The long-range ferromagnetic order is preserved in the samples right up to $\gamma \approx 0.05$; this is indicated by the identical behavior of the ZFC and FC curves. The Curie temperature for the anion-deficient solid solution $La_{0.70}Sr_{0.30}MnO_{2.95}$ is about 282 K (Fig. 3c). The ZFC and FC curves exhibit different behavior for samples with $\gamma > 0.05$. In all probability, this is an indication of the absence of long-range ferromagnetic order. The FC curve for $La_{0.70}Sr_{0.30}MnO_{2.87}$ (Fig. 3b) is concave and descends gradually at a temperature higher than approximately 49 K. The transition to the paramagnetic state is quite broad. The ZFC curve exhibits a peak at $T_f \sim 49$ K. A similar behavior is observed for the $La_{0.70}Sr_{0.30}MnO_{2.80}$ sample (Fig 3a). The peak of the ZFC curve is slightly lowered and is observed at a temperature of about 43 K. Thus, anion-deficient samples with $\gamma \geq 0.13$ exhibit properties typical of cluster spin glasses. This is also confirmed by a very smooth variation of the temperature derivative $\partial M_{FC}/\partial T$ of the FC curve (see inset to Fig. 3). A gradual decrease in the value of T_f with increasing γ indicates a decrease in the size of ferromagnetic clusters. The $La_{0.70}Sr_{0.30}MnO_{2.75}$ sample exhibits a behavior typical of an antiferromagnet with $T_N = 124$ K (Fig. 4).

Figure 5 shows the field dependence of the atomic magnetic moment for samples of the anion-deficient series $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$. A few salient features can be singled out on these curves. First, the magnetic moment for samples with $\gamma > 0.05$ is not saturated completely in fields up to 16 kOe, which slightly complicates the determination of the spontaneous atomic magnetic moment M_s. Second, M_s does not attain the theoretically attainable value of $[(0.70 + 2\gamma) \times 4 + (0.30 - 2\gamma) \times 3]\mu_B$ per structural unit for any anion-deficient sample under the assumption of complete ferromagnetic ordering of the Mn^{3+} and Mn^{4+} ions. The value of M_s decreases continuously with increasing γ, although it should increase according to the theory. A noticeable field hysteresis is observed for samples of $La_{0.70}Sr_{0.30}MnO_{2.87}$ (Fig. 5b) and $La_{0.70}Sr_{0.30}MnO_{2.80}$ (Fig. 5c). These samples are also characterized by a substantial residual magnetization and coercive force.

Figures 6 and 7 show the generalized results of magnetic investigations of anion-deficient samples of $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$. As the value of γ increases, a steady decrease is observed in the spontaneous atomic magnetic moment from $M_s \approx 3.66\mu_B$ per structural unit ($\gamma = 0$) to $0.02\mu_B$ per structural unit ($\gamma = 0.25$). The sharpest decrease in the value of M_s is observed for the sample with $\gamma = 0.13$. However, M_s does not attain its theoretical value for any composition. A considerable increase in the coercive force was observed for $\gamma \geq 0.13$.

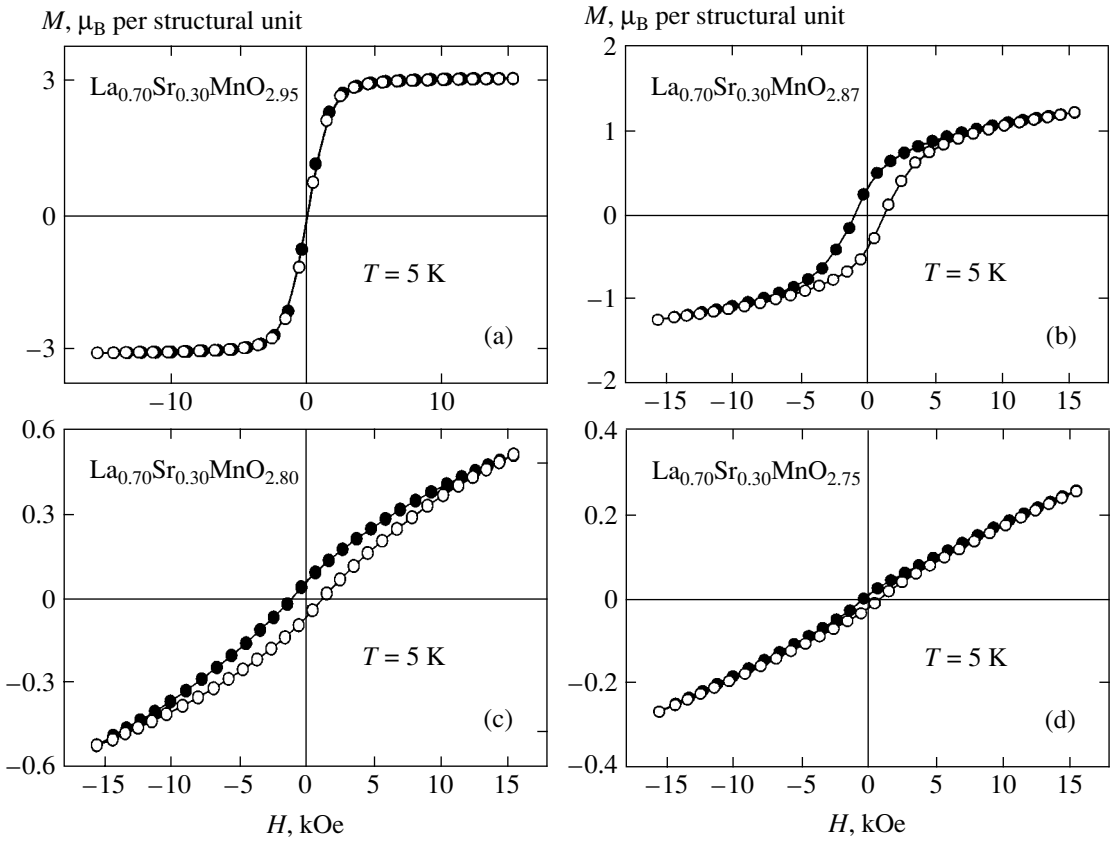


Fig. 5. Magnetic field dependence of the atomic magnetic moment for anion-deficient solid solutions of (a) $La_{0.70}Sr_{0.30}MnO_{2.95}$, (b) $La_{0.70}Sr_{0.30}MnO_{2.87}$, (c) $La_{0.70}Sr_{0.30}MnO_{2.80}$, and (d) $La_{0.70}Sr_{0.30}MnO_{2.75}$ at 5 K. The dark symbols indicate the results of measurements in a decreasing field and light symbols, in an increasing field.

This fact indicates that the magnetic anisotropy considerably increases with oxygen deficiency.

Figure 7a shows the effect of oxygen vacancies on the magnetic properties of Sr-substituted manganites. The figure shows the linear interpolation of the critical temperature of magnetic phase transition to the paramagnetic state depending on the concentration of Mn^{3+} ions for two series of samples: (i) stoichiometric $La_{1-x}Sr_xMnO_3$ and (ii) anion-deficient $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$. These two series have the same concentration of Mn^{3+} ions under the condition $x = 0.30 - 2\gamma$. It can be clearly seen that these two series of samples reveal identical behavior of T_{mo} upon an increase in the concentration of Mn^{3+}. The value of this temperature decreases, but the decrease is not the same in both series. For the anion-deficient series, the decrease is sharper. This leads to the conclusion that oxygen vacancies are an additional parameter making a significant contribution to the formation of magnetic properties of systems with the same chemical composition but with different valences of manganese ions. The presence of oxygen vacancies in the region with a high concentration of Mn^{3+} ions suppresses the ferromagnetic properties the more strongly, the larger the ratio of concentrations of Mn^{3+} and Mn^{4+} ions. The coefficient describing the effect of oxygen nonstoichiometry on the decrease in the value of T_{mo} can be obtained from the difference in temperatures $T_{mo}(x)$

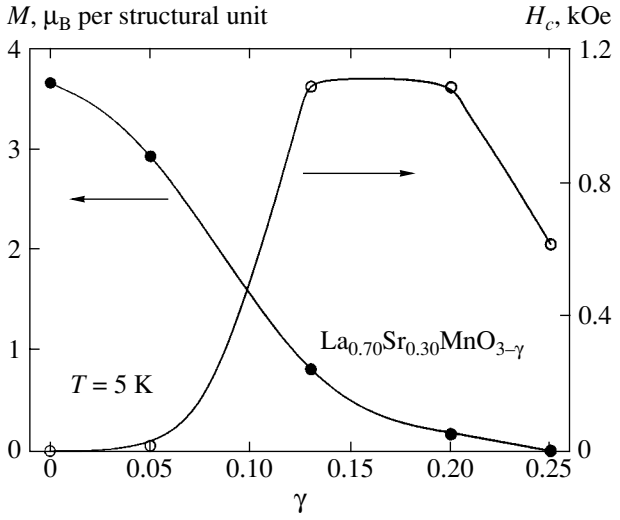


Fig. 6. Concentration dependence of the values of spontaneous atomic magnetic moment (dark symbols) and coercive force (light symbols) for a number of anion-deficient solid solutions $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ ($0 \le \gamma \le 0.25$) at 5 K.

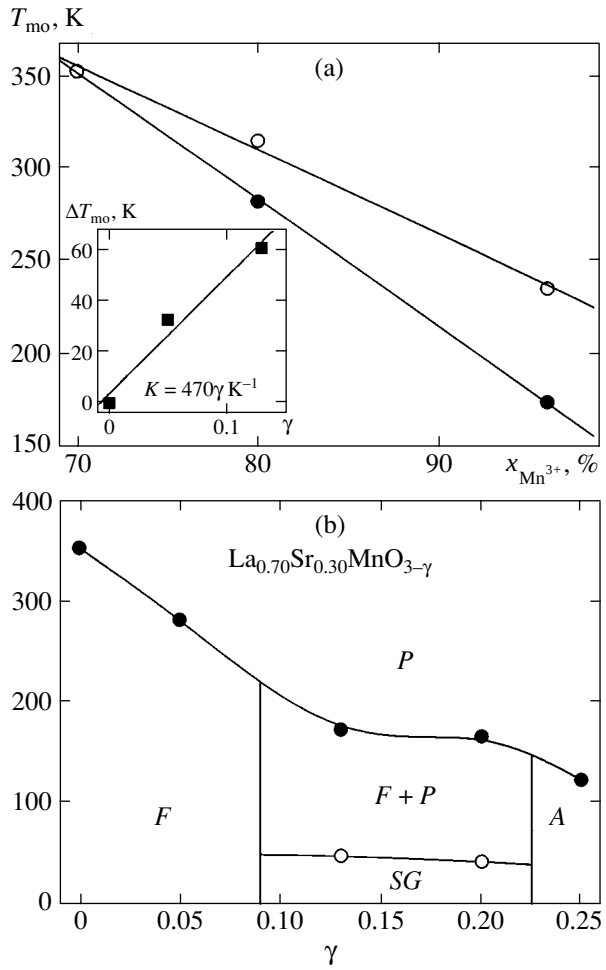


Fig. 7. (a) Critical temperature of the magnetic phase transition as a function of (a) concentration of the Mn^{3+} ions for solid solutions of the stoichiometric composition $La_{1-x}Sr_xMnO_3$ (light symbols) and anion-deficient solid solutions $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ (dark symbols) and (b) magnetic phase diagram for the system of $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ anion-deficient solid solutions. *F* indicates a ferromagnet, *F* + *P* stands for the mixed magnetic state formed by ferromagnetic clusters in the paramagnetic matrix, *SG* denotes a cluster spin glass, *A* indicates an antiferromagnet and *P*, a paramagnet. Solid symbols correspond to the temperature of a transition to the paramagnetic state, while light symbols correspond to the temperature of freezing of magnetic moments in ferromagnetic clusters. The inset to Fig. 7a demonstrates the difference in the critical temperatures of the magnetic transition for the stoichiometric and anion-deficient series as a function of the oxygen vacancy concentration.

and $T_{mo}(\gamma)$ for the stoichiometric and anion-deficient series, respectively, which is defined as

$$\Delta T_{mo}(\gamma) = T_{mo}(x = 0.30 - 2\gamma) - T_{mo}(\gamma). \quad (2)$$

This coefficient is $K \sim 470\gamma K^{-1}$.

The substituted lanthanum manganites $La_{1-x}Sr_xMnO_{3-\gamma}$ with an oxygen deficiency are interesting for experimental studies of the indirect 180° exchange between manganese ions [30]. It should be noted that La^{3+}, Sr^{2+}, and O^{2-} ions are diamagnetic. This fact considerably simplifies the interpretation of the results of magnetic studies.

Zener worked out a special theory of indirect exchange interactions via charge carriers (double exchange) to explain the magnetic properties of substituted manganites [31, 32]. This type of dependence is observed preferably for ferromagnets with metal-type conductivity, but several facts cannot be explained by the double-exchange theory [33].

The basic tenets of the theory of superexchange, a specific type of indirect exchange interaction, were formulated by Goodenough. He assumed that ferromagnetism is due not only to a strong double exchange, but also to the peculiarities of exchange interactions in the system of Jahn–Teller Mn^{3+} ions. In the absence of static Jahn–Teller distortions, the orbital configuration of 3*d* electrons is determined by the position of the nuclei of manganese ions. In other words, a correlation must be observed between the electron configuration and vibrational modes of the nucleus so that the interaction Mn^{3+}–O–Mn^{3+} becomes ferromagnetic. This assumption, known as the Goodenough quasistatic hypothesis and based on virtual exchange of electrons between manganese ions, predicts the anisotropic nature of their exchange interactions [34].

In the general case, the energy of the exchange interaction in the system of magnetic ions is defined by the Heisenberg Hamiltonian [35]

$$\hat{H}_{ex} = -\sum_{i,j} J_{i,j}\hat{\mathbf{S}}_i\hat{\mathbf{S}}_j, \quad (3)$$

where $\hat{\mathbf{S}}_i$ and $\hat{\mathbf{S}}_j$ are the operators of the total spins of the *i*th and *j*th magnetic ions, respectively, and $J_{i,j}$ is the exchange integral defined as

$$J_{i,j} = \int \varphi_i^*(1)\varphi_j^*(2)\frac{e_0^2}{|\mathbf{r}_1 - \mathbf{r}_2|}\varphi_j(1)\varphi_i(2)d\mathbf{r}_1 d\mathbf{r}_2, \quad (4)$$

where $\varphi_i(1)$ and $\varphi_j(2)$ are the wavefunctions of the *i*th and *j*th ions in states 1 and 2, respectively, e_0 is the electron charge, and $\mathbf{r}_1$ and $\mathbf{r}_2$ are the radius vectors of states 1 and 2. A positive exchange integral ($J_{i,j} > 0$) in Eq. (4) corresponds to the ferromagnetic interaction since the exchange energy minimum of the system is observed for $\mathbf{S}_i \uparrow\uparrow \mathbf{S}_j$. For $J_{i,j} < 0$, antiferromagnetic ordering of spins sets in.

The calculations of exchange integrals between the total spins of manganese ions are extremely tedious and do not always lead to a value close to the one observed. Consequently, it is simpler and more constructive to seek possible magnetic phase states by writing the

exchange Hamiltonian of the system proceeding from physical considerations and experimental results, and treating various quantities being measured as phenomenological parameters. In this case, following Goodenough [30], it is assumed that $J_{i,j}^{2,2}$ (in the absence of orbital ordering) and $J_{i,j}^{3/2,3/2}$ have positive signs, while $J_{i,j}^{2,3/2}$ have negative signs (the superscripts correspond to total spins $S = 2$ and 3/2 for the Mn^{3+} and Mn^{4+} ions, respectively).

It is well known [36] that the superexchange interactions Mn^{3+}–O–Mn^{3+} are anisotropic in the orbitally ordered phase. For (001) planes, the exchange integral is greater than zero ($J_{i,j}^{2,2}(001) > 0$), while along the [001] direction, this integral is smaller than zero ($J_{i,j}^{2,2}[001] < 0$). In the orbitally disordered phase, these interactions are isotropic, i.e., positive along all directions ($J_{i,j}^{2,2} > 0$). The stoichiometric solid solution $La_{0.70}Sr_{0.30}MnO_3$ is not orbitally ordered; consequently, the superexchange interactions Mn^{3+}–O–Mn^{3+} are positive.

The emergence of oxygen vacancies reverses the sign of the exchange integral for the interactions Mn^{3+}–O–Mn^{3+}. The superexchange interaction $Mn^{3+}(5)$–O–$Mn^{3+}(5)$ for the pentahedral coordination of manganese ions is antiferromagnetic. This is confirmed by the compound $Ca^{2+}Mn^{3+}O_{2.50}^{2-}$ which is antiferromagnetic [37, 38]. The $Mn^{3+}(5)$ ions in this compound are in the pentahedral coordination of oxygen anions. Thus, anion-deficient samples of $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ are ferromagnets with values of T_{mo} and M_s decreasing slightly upon an increase in γ in the interval $0 \le \gamma \le 0.05$.

For $\gamma > 0.05$, the volume of the antiferromagnetic orbitally disordered phase associated with the interaction $Mn^{3+}(5)$–O–$Mn^{3+}(5)$ begins to increase substantially, leading to a decrease in the spontaneous magnetic moment. The competition in the interaction between antiferromagnetically and ferromagnetically ordered clusters probably leads to the cluster spin glass state with gradually decreasing freezing temperature for the magnetic moments of ferromagnetic clusters. The orientation of the magnetic moments of clusters in a spin glass in the temperature range below T_f does not exhibit any spatial periodicity. This orientation changes in space at random like the arrangement of atoms in an ordinary glass. In contrast to paramagnets, in which magnetic moments fluctuate in time, spin glasses are characterized by "frozen" magnetic moments, i.e., by nonzero time-averaged vector quantities. A state of the spin glass type is often observed in heterogeneous magnetic systems such as the alloys Co–Cu and Co–Ag [39, 40]. In such alloys, ferromagnetic grains are implanted in a nonferromagnetic matrix. The formation of the spin glass state for the anion-deficient solid solutions $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ is confirmed by the clearly manifested peak on the temperature dependence described by the ZFC curve. A gradual decrease in T_f indicates a decrease in the size of ferromagnetically ordered clusters.

In this study, we obtained a quantitative estimate of the mean size of ferromagnetic clusters formed upon an increase in the number of oxygen vacancies. To do this, we used the Bean–Livingstone formula [41] connecting the mean size of ferromagnetic inclusions in a dia- or a paramagnetic matrix with the magnetic crystallographic anisotropy constant (which is precisely equal to the volume energy density of magnetic crystallographic anisotropy) and the critical temperature T_f,

$$\langle K \rangle \langle V \rangle = k_B T_f, \tag{5}$$

where $\langle K \rangle$ is the mean value of the magnetic crystallographic anisotropy constant for a ferromagnetic cluster, $\langle V \rangle$ is its mean volume, k_B is the Boltzmann constant, and T_f is the freezing temperature (corresponding to the peak on the ZFC curve). The anisotropy constants can be obtained from the equation [42] for the magnetic crystallographic anisotropy energy in the (100) plane,

$$E_a = K_1 \sin^2\theta + K_2 \sin^4\theta, \tag{6}$$

where K_1 and K_2 are the magnetic crystallographic anisotropy constants and θ is the angle between the magnetization and the [010] axis. The magnetic crystallographic anisotropy energy is determined by the area between the $M(H)$ curves measured for single crystals along the [010] and [001] directions.

In [43], it was found that $K_1 = 45.3 \times 10^4$ erg cm^{-3} and $K_2 = 73.6 \times 10^4$ erg cm^{-3} for a $La_{0.744}Ba_{0.186}MnO_3$ manganite single crystal at 4.2 K. The mean size of a ferromagnetic cluster for samples of $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ ($\gamma \ge 0.13$) with $T_F \sim 49$ K has been approximately estimated at 50 nm (for the above data). Thus, we can state that ferromagnetic clusters of a nanometric size are formed as a result of the concentration phase transition to the spin glass.

When the concentration of oxygen vacancies is $\gamma = 0.25$, a sample of $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ becomes a charge-disordered antiferromagnet. This sample contains 20% of Mn^{2+} ions. As a rule, Mn^{3+} and Mn^{2+} ions in manganites are ordered so that they are antiparallel [44]. Since these ions have different magnetic moments ($4\mu_B$ and $5\mu_B$ per structural unit), this ensures a ferromagnetic contribution amounting to $1\mu_B$ per structural unit. The observed value of the moment equal approximately to $0.02\mu_B$ per structural unit indicates that Mn^{2+} ions are probably eliminated by oxygen vacancies from the exchange interactions. The vacancies must be predom-

inantly located near the Mn^{2+} due to the weakness of electrostatic repulsion.

4. CONCLUSIONS

In this study, we have reported on the results of experimental investigation of the chemical phase composition, crystal structure, and magnetic properties of a manganite with the perovskite structure $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ ($0 \le \gamma \le 0.25$) for various concentrations of oxygen vacancies. The stoichiometric solid solution $La_{0.70}Sr_{0.30}MnO_3$ was obtained in air by conventional ceramic technology, while the anion-deficient compounds $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ were obtained in vacuum using the topotactic reaction technique. The microscopic structure and the chemical phase composition of all the samples were studied using a scanning electron microscope. It was found that the mean grain size in the stoichiometric solid solution $La_{0.70}Sr_{0.30}MnO_3$ obtained at 1550°C is about 10 μm, while its value for $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ anion-deficient solid solutions is 5 μm. Thermal treatment at lower temperatures leads to a decrease in the mean grain size. It was found that samples with $0 \le \gamma \le 0.13$ possess a rhombohedral unit cell (space group $R\bar{3}c$, $Z = 2$), while samples with $\gamma \ge 0.20$ have a tetragonal unit cell (space group $I4/mcm$, $Z = 2$). Experimental evidence is provided for the variation of the magnetic phase state of the investigated manganite upon a decrease in the oxygen concentration. It is shown that $La_{0.70}Sr_{0.30}MnO_{3-\gamma}$ anion-deficient solid solutions experience a series of successive magnetic phase transformations in the ground state from a ferromagnet ($0 \le \gamma \le 0.05$) to a charge-disordered antiferromagnet ($\gamma = 0.25$) via an inhomogeneous magnetic state similar to a cluster spin glass ($0.13 \le \gamma \le 0.20$). The mean size of ferromagnetic clusters in the spin glass state was estimated as $r \approx 50$ nm. It is shown that oxygen vacancies make a substantial contribution to the properties of manganites. The generalized magnetic data are presented as concentration dependences of the spontaneous magnetic moment and the coercive force. The magnetic phase diagram has been constructed. The most probable mechanism of formation of the magnetic phase state in Sr-substituted anion-deficient manganites is considered. It is assumed that in the absence of orbital ordering, a decrease in the coordination number of a magnetic ion leads to the sign reversal of indirect superexchange interactions Mn^{3+}–O–Mn^{3+}.

ACKNOWLEDGMENTS

This study was partly supported by the State program on Basic Research "Nanomaterials and Nanotechnologies" (task no. 3.3) and by the Belarus Republic Foundation for Basic Research (grant no. F04R-076).

REFERENCES

1. K. Chabara, T. Ohno, M. Kasai, and Y. Kozono, Appl. Phys. Lett. **63**, 1990 (1993).
2. R. Helmolt, J. Wecker, B. Holzapfel, *et al.*, Phys. Rev. Lett. **71**, 2331 (1993).
3. P. Schiffer, A. P. Ramirez, W. Bao, and S. W. Cheong, Phys. Rev. Lett. **75**, 3336 (1995).
4. P. G. Radaelli, D. E. Cox, M. Marezio, *et al.*, Phys. Rev. Lett. **75**, 4488 (1995).
5. J. M. D. Coey, M. Viret, and S. von Molnar, Adv. Phys. **48**, 167 (1999).
6. A. Urushibara, Y. Moritomo, T. Arima, *et al.*, Phys. Rev. B **51**, 14103 (1995).
7. S. V. Trukhanov, J. Mater. Chem. **13**, 347 (2003).
8. J. Töpfer and J. B. Goodenough, J. Solid State Chem. **130**, 117 (1997).
9. R. A. DeSouza, M. S. Islamb, and E. Ivers-Tiffée, J. Mater. Chem. **9**, 1621 (1999).
10. Y. Tokura and Y. Tomioka, J. Magn. Magn. Mater. **200**, 1 (1999).
11. E. Dagotto, T. Hotta, A. Moreo, *et al.*, Phys. Rep. **344**, 1 (2001).
12. S. M. Dunaevskiĭ, Fiz. Tverd. Tela (St. Petersburg) **46**, 193 (2004) [Phys. Solid State **46**, 193 (2004)].
13. G. Matsumoto, J. Phys. Soc. Jpn. **29**, 606 (1970).
14. G. Matsumoto, J. Phys. Soc. Jpn. **29**, 615 (1970).
15. I. Dzialoshinsky, J. Phys. Chem. Solids **4**, 241 (1958).
16. T. Moriya, Phys. Rev. **120**, 91 (1960).
17. S. V. Trukhanov, N. V. Kasper, I. O. Troyanchuk, *et al.*, J. Solid State Chem. **169**, 85 (2002).
18. A. Barnabé, F. Millange, A. Maignan, *et al.*, Chem. Mater. **10**, 252 (1998).
19. A. Barnabé, M. Gaudon, C. Bernard, *et al.*, Mater. Res. Bull. **39**, 725 (2004).
20. B. C. Tofield and W. R. Scott, J. Solid State Chem. **100**, 183 (1974).
21. C. R. Wiebe, J. E. Greedan, J. S. Gardner, *et al.*, Phys. Rev. B **64**, 064421 (2001).
22. S. V. Trukhanov, M. V. Bushinskiĭ, I. O. Troyanchuk, and H. Szymczak, Zh. Éksp. Teor. Fiz. **126**, 874 (2004) [JETP **99**, 756 (2004)].
23. N. Abdelmoula, K. Guidara, A. Cheikh-Rouhou, and E. Dhari, J. Solid State Chem. **151**, 139 (2000).
24. S. V. Trukhanov, I. O. Troyanchuk, N. V. Pushkarev, and H. Szymczak, Zh. Éksp. Teor. Fiz. **122**, 356 (2002) [JETP **95**, 308 (2002)].
25. Y. M. Goldsmidt, J. Maten Naturwid. Kl. **2**, 97 (1926).
26. E. L. Nagaev, Phys. Rep. **346**, 387 (2001).
27. J. B. Goodenough, Phys. Rev. **100**, 564 (1955).
28. I. O. Troyanchuk, S. V. Trukhanov, H. Szymczak, *et al.*, Zh. Éksp. Teor. Fiz. **120**, 183 (2001) [JETP **93**, 161 (2001)].
29. R. D. Shannon, Acta Cryst. A **32**, 751 (1976).

30. J. B. Goodenough, *Magnetism and the Chemical Bond* (Interscience, New York, 1963; Metallurgiya, Moscow, 1968).
31. C. Zener, Phys. Rev. **82**, 403 (1951).
32. C. Zener, Phys. Rev. **82**, 440 (1951).
33. S. V. Trukhanov, N. V. Kasper, I. O. Troyanchuk, *et al.*, Phys. Status Solidi B **233**, 321 (2002).
34. J. B. Goodenough, A. Wold, R. J. Arnott, and N. Menyuk, Phys. Rev. **124**, 373 (1961).
35. C. Kittel, *Introduction to Solid State Physics*, 7th ed. (Wiley, New York, 1996; Nauka, Moscow, 1978).
36. S. V. Trukhanov, I. O. Troyanchuk, M. Hervieu, *et al.*, Phys. Rev. B **66**, 184424 (2002).
37. K. R. Poeppelmeier, M. E. Leonowicz, and J. M. Longo, J. Solid State Chem. **44**, 89 (1982).
38. K. R. Poeppelmeier, M. E. Leonowicz, J. C. Scanlon, *et al.*, J. Solid State Chem. **45**, 71 (1982).
39. S. Nafis, J. A. Woollam, Z. S. Shan, and D. J. Sellmyer, J. Appl. Phys. **70**, 6050 (1991).
40. F. Conde, C. Gomez-Polo, and A. Hernando, J. Magn. Magn. Mater. **138**, 123 (1994).
41. C. P. Bean and J. D. Livingstone, J. Appl. Phys. **30**, S120 (1959).
42. H. J. Williams, Phys. Rev. **52**, 747 (1937).
43. Yu. Bukhantsev, Ya. M. Mukovskii, and H. Szymczak, J. Magn. Magn. Mater. **272**, 2053 (2004).
44. I. O. Troyanchuk, S. V. Trukhanov, D. D. Khalyavin, and H. Szymczak, J. Magn. Magn. Mater. **208**, 217 (2000).

Translated by N. Wadhwa