Journal of Experimental and Theoretical Physics, Vol. 99, No. 2, 2004, pp. 225–233.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 2, 2004, pp. 259–267.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Vector Meson Dominance Pion Electromagnetic Form Factor with σ-Model Loop Corrections¶

S. K. Dubinsky[a], A. Yu. Korchin[b], and N. P. Merenkov[b]

[a]*Kharkov National University, Kharkov, 61077 Ukraine*
[b]*National Science Center Kharkov Institute of Physics and Technology, Kharkov, 61108 Ukraine*
e-mail: korchin@kipt.kharkov.ua, merenkov@kipt.kharkov.ua

Received December 18, 2003

Abstract—A model is developed for the electromagnetic form factor of the pion. One-loop corrections are included in the linear σ-model. The ρ-meson contribution is added in an extended VMD model. The form factor, calculated without fitting parameters, is in a good agreement with experiment for spacelike and timelike photon momenta. Loop corrections to the two-pion hadronic contribution $a_\mu^{(had,\pi)}$ to the muon anomalous magnetic moment are calculated. The optimal value of the σ-meson mass appears to be very close to the ρ-meson mass.

1. INTRODUCTION

It has recently been understood that the pion electromagnetic form factor is a very important physical quantity that plays a key role in testing the Standard Model at the electroweak precision level. The reason is that, at low energies, the production cross section

$$\sigma(e^+e^- \to \pi^+\pi^-) = \frac{\pi\alpha^2}{3s}\left(1-\frac{4m_\pi^2}{s}\right)^{3/2}|F_\pi(s)|^2, \quad (1)$$

where s is the squared total energy in center-of-mass system, α is the fine structure constant, and m_π is the pion mass, dominates over the other hadronic channels and accounts for more than 70% of the total hadronic contribution to the muon anomalous magnetic moment (AMM)

$$a_\mu = \frac{g_\mu - 2}{2}.$$

The recent measurement of a_μ from the Brookhaven E821 experiment [1] has boosted interest in a renewed theoretical calculation of this quantity [2].

The main ingredient of the theoretical prediction of a_μ, which is responsible for the bulk of the theoretical error, is the hadronic contribution to the vacuum polarization. The contribution of the $\pi^+\pi^-$ channel to the electron–positron annihilation process can be written in terms of the form factor $F_\pi(s)$ via the dispersion integral [3]

$$a_\mu^{(had,\pi)} = \frac{1}{4}\int\limits_{4m_\pi^2}^{\infty} K(s)\left(1-\frac{4m_\pi^2}{s}\right)^{3/2}|F_\pi(s)|^2 ds,$$
$$K(s) = \frac{\alpha^2}{3\pi^2 s}\int\limits_0^1 \frac{x^2(1-x)}{x^2+(1-x)s/m_\mu^2}dx, \quad (2)$$

where m_μ is the muon mass.

Conventional strategy of the model-independent evaluation of this integral consists in using precise experimental data (at least at low energies, where perturbative QCD cannot reliably be applied). However, the predicted accuracy, which should be reached soon in the E821 experiment, requires calculation of electromagnetic radiative corrections to cross section (1) [4]. Apart from pure $\pi^+\pi^-$ events, electromagnetic radiative corrections include the $\pi^+\pi^-\gamma$ process where the photon is radiated from the final pions. In current experiments at Φ and B factories, based on the radiative return method [5], this contribution cannot be extracted in a model-independent way[1] and the corresponding procedure requires model-dependent approaches. This, in turn, stimulates development and study of different theoretical models of pion–photon interaction. The simplest is the pointlike scalar quantum electrodynamics (sQED) [7] combined with the standard vector meson dominance (VMD) model (see, e.g., [8]) for description

¶This article was submitted by the authors in English.

[1] Even in direct scanning experiments, a model-independent treatment of $\pi^+\pi^-\gamma$ events suggested in [6] seems too complicated to be used in the near future.

of the $\gamma^* \longrightarrow \pi^+\pi^-$ transition form factor in the ρ-resonance region. Such a model was used in [4] for construction of the Monte Carlo event generator.

In the present paper, we consider a modification of the pion electromagnetic form factor in the linear σ-model [9] with spontaneously broken chiral symmetry, which includes the nucleon sector. The ρ-meson contribution is added following [10, 11]. In particular, the ρ coupling to the pion and nucleon is introduced through gauge-covariant derivatives, while the direct γρ coupling has an explicitly gauge-invariant form. We calculate the pion form factor in the one-loop approximation in the strong interaction and compare $F_\pi(s)$ with precise data from elastic $e^-\pi^+$ scattering and e^+e^- annihilation in the pion pair.

We take the loop corrections to $a_\mu^{(had,\pi)}$ into account. In general, because the σ-model Lagrangian contains the sQED Lagrangian as a constituent part, one can expect that the difference between the predictions of σ-model + VMD and sQED + VMD is small. Indeed, it follows from our calculation that the loop corrections increase the low-energy part of the right-hand side of Eq. (2) by about 2%, as compared with sQED + VMD.

2. FORMALISM

2.1. Lagrangian

The Lagrangian of the model consists of two parts,

$$\mathcal{L} = \mathcal{L}^{(1)} + \mathcal{L}^{(2)}.$$

The first is the Lagrangian of the chiral linear σ-model [9] with an explicit symmetry-breaking term $c\phi$. After spontaneous breaking of chiral symmetry and redefinition of the scalar field via

$$\phi = \sigma + v,$$

where $v = \langle\phi\rangle$ is the vacuum expectation value, the Lagrangian takes the form

$$\begin{aligned}\mathcal{L}^{(1)} &= \bar N(i\partial\!\!\!/ - m_N)N + \frac{1}{2}[(\partial\sigma)^2 - m_\sigma^2\sigma^2] \\ &+ \frac{1}{2}[(\partial\vec\pi)^2 - m_\pi^2\vec\pi^2] \\ &- g_\pi\bar N(\sigma + i\gamma_5\vec\tau\vec\pi)N - \lambda(\sigma^2 + \vec\pi^2) \\ &\times\left[v\sigma + \frac{1}{4}(\sigma^2 + \vec\pi^2)\right] + \text{const},\end{aligned} \tag{3}$$

where N, $\vec\pi$, and σ are the respective fields of the nucleon, pion, and meson with vacuum quantum numbers; g_π is the coupling constant; λ is the parameter of the meson potential; and

$$\partial\!\!\!/ \equiv \partial^\mu\gamma_\mu, \quad (\partial\sigma)^2 = \partial^\mu\sigma\partial_\mu\sigma,$$

etc. All parameters of the model are related via

$$m_N = g_\pi v, \quad m_\sigma^2 = 2\lambda v^2 + m_\pi^2, \quad m_\pi^2 = \frac{c}{v}. \tag{4}$$

Moreover, in the tree-level approximation, $v = f_\pi$, where $f_\pi = 93.2$ MeV is the pion weak decay constant. More details on the σ-model can be found, e.g., in [12, Ch. 5, Sec. 2.6].

The second part of the Lagrangian includes coupling to electromagnetic field A^μ and field ρ^μ of the *neutral* ρ-meson. This coupling can be obtained using the minimal substitutions

$$\begin{aligned}\partial^\mu N &\longrightarrow \left(\partial^\mu + ie\frac{1+\tau_3}{2}A^\mu + ig_\rho\frac{\tau_3}{2}\rho^\mu\right)N, \\ \partial^\mu\pi^a &\longrightarrow \partial^\mu\pi^a + (eA^\mu + g_\rho\rho^\mu)\varepsilon^{3ab}\pi^b, \\ &a, b = 1, 2, 3, \\ \partial^\mu\sigma &\longrightarrow \partial^\mu\sigma,\end{aligned} \tag{5}$$

where e is the proton charge, g_ρ is the coupling constant, and τ_3 is the third component of the Pauli vector $\vec\tau = (\tau_1, \tau_2, \tau_3)$. In addition, we include the direct coupling of the photon to the ρ-meson in the version of VMD model from [10, 11]. We thus obtain

$$\begin{aligned}\mathcal{L}^{(2)} &= \frac{1}{2}m_\rho^2\rho_\mu\rho^\mu - \frac{1}{4}\rho_{\mu\nu}\rho^{\mu\nu} - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \\ &- (eA_\mu + g_\rho\rho_\mu)(\vec\pi\times\partial^\mu\vec\pi)_3 + (eA_\mu + g_\rho\rho_\mu)^2(\vec\pi^2 - \pi_3^2) \\ &- g_\rho\bar N\gamma^\mu\frac{\tau_3}{2}N\rho_\mu - e\bar N\gamma^\mu\frac{1+\tau_3}{2}NA_\mu - \frac{e}{2f_\rho}\rho_{\mu\nu}F^{\mu\nu},\end{aligned} \tag{6}$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \quad \rho_{\mu\nu} = \partial_\mu\rho_\nu - \partial_\nu\rho_\mu,$$

and f_ρ determines the γρ coupling. In Eq. (6), we assume equal coupling constants of ρ to the pion and the nucleon in accordance with the universality hypothesis of Sakurai (see, e.g., [12, Ch. 5, Sec. 4]). At the same time, the γρ coupling constant f_ρ does not necessarily coincide with g_ρ. Lagrangian (6) is gauge-invariant because of the form of the γρ coupling. We mention that the nucleon contribution is also included in Lagrangian (6), contrary to [10, 11].

2.2. Counterterms and Renormalization

Because one of the purposes of the present paper is to take loop corrections to the pion electromagnetic vertex into account, we need to specify the way in which the parameters are renormalized. We use the conventional approach and assume that Lagrangians (3) and (6) involve the "bare" fields, coupling constants and masses, to be marked by the subscript 0. The bare fields require rescaling,

$$(\vec{\pi}_0, \sigma_0) = \sqrt{Z_\pi}(\vec{\pi}, \sigma), \quad N_0 = \sqrt{Z_N}N,$$

$$\rho_0^\mu = \sqrt{Z_\rho}\rho^\mu, \quad A_0^\mu = \sqrt{Z_A}A^\mu,$$

where Z_π, Z_N, Z_ρ, and Z_A are the respective wave-function renormalization constants for the pion (or sigma), nucleon, rho, and photon. The procedure for obtaining the counterterm Lagrangian is known (see, e.g., [13, Ch. 10]). For $\mathscr{L}^{(1)}$, the corresponding counterterm Lagrangian is given by

$$\begin{aligned}\mathscr{L}_{ct}^{(1)} &= \delta_{Z_N}\bar{N}i\partial\!\!\!/N - \delta_{g_\pi}v\bar{N}N - \delta_{g_\pi}\bar{N}(\sigma + i\gamma_5\vec{\tau}\vec{\pi})N \\ &+ \frac{1}{2}\delta_{Z_\pi}[(\partial\vec{\pi})^2 + (\partial\sigma)^2] \\ &- \frac{1}{2}(\delta_\mu + 3\delta_\lambda v^2)\sigma^2 - \frac{1}{2}(\delta_\mu + \delta_\lambda v^2)\vec{\pi}^2 \\ &- \frac{1}{4}\delta_\lambda(\vec{\pi}^2 + \sigma^2)^2 - \delta_\lambda v(\sigma\vec{\pi}^2 + \sigma^3) \\ &- [(\delta_\mu + \delta_\lambda v^2)v - \delta_c]\sigma + \text{const}.\end{aligned} \tag{7}$$

There are six constants δ_{Z_π}, δ_{Z_N}, δ_μ, δ_λ, δ_{g_π}, δ_c, which can be fixed by imposing six conditions on the Green functions in general. In calculating the pion electromagnetic vertex, only one constant δ_{Z_π} is needed (see Section 2.3).

For Lagrangian (6), we can first define the physical values of the electric charge

$$e = e_0\sqrt{Z_A}$$

and the ρ coupling

$$g_\rho = g_{\rho 0}\sqrt{Z_\rho}.$$

It is also convenient to introduce

$$\hat{m}_\rho = \sqrt{Z_\rho}m_{\rho 0}$$

(the ρ-meson mass in the absence of coupling to pions). The counterterm Lagrangian can be written as

$$\begin{aligned}\mathscr{L}_{ct}^{(2)} &= -\delta_{Z_A}\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \delta_{Z_\rho}\frac{1}{4}\rho_{\mu\nu}\rho^{\mu\nu} \\ &- \delta_{Z_\pi}(eA_\mu + g_\rho\rho_\mu)(\vec{\pi}\times\partial^\mu\vec{\pi})_3 \\ &+ \delta_{Z_\pi}(eA_\mu + g_\rho\rho_\mu)^2(\vec{\pi}^2 - \pi_3^2) \\ &- \delta_{Z_N}g_\rho\bar{N}\gamma^\mu\frac{\tau_3}{2}N\rho_\mu - \delta_{Z_N}e\bar{N}\gamma^\mu\frac{1+\tau_3}{2}NA_\mu - \delta_{f_\rho}\frac{e}{2}\rho_{\mu\nu}F^{\mu\nu}.\end{aligned} \tag{8}$$

It follows that we in general need three additional constants δ_{Z_A}, δ_{Z_ρ}, and δ_{f_ρ} once δ_{Z_π} and δ_{Z_N} are fixed. Finally, the total Lagrangian is the sum[2]

$$\mathscr{L} = \mathscr{L}^{(1)} + \mathscr{L}^{(2)} + \mathscr{L}_{ct}^{(1)} + \mathscr{L}_{ct}^{(2)}. \tag{9}$$

2.3. Contribution to the Pion Electromagnetic Form Factor from the σ-Model Sector

The Feynman rules for Lagrangian (9) are obtained according to the standard prescriptions [14]. The counterterm constants can be found by imposing the following constraints on the respective self-energy operators of the pion, sigma-meson, and nucleon:

$$\begin{aligned}\Sigma_\pi(m_\pi^2) &= \frac{d}{dp^2}\Sigma_\pi(p^2)\Big|_{p^2=m_\pi^2} = \Sigma_\sigma(m_\sigma^2) = 0, \\ \Sigma_N(p\!\!\!/)\big|_{p\!\!\!/=m_N} &= \frac{d}{dp\!\!\!/}\Sigma_N(p\!\!\!/)\Big|_{p\!\!\!/=m_N} = 0.\end{aligned} \tag{10}$$

These conditions imply that the respective pole positions of the pion, nucleon, and sigma propagators are located at the physical mass of the pion, nucleon, and sigma. In addition, the residue of the pion and nucleon propagators is unity, ensuring the absence of renormalization for the external pion and nucleon (but not for the external sigma-meson). We also impose the condition $\langle\sigma\rangle = 0$ by requiring that the so-called tadpole diagrams vanish. Accordingly, the tadpole diagrams do not contribute to the quantities calculated below.

In calculation of the loop integrals, we use the dimensional regularization method (see, e.g., [13,

[2] The mass m_ρ in $\mathscr{L}^{(2)}$ is replaced by $\hat{m}_\rho$.

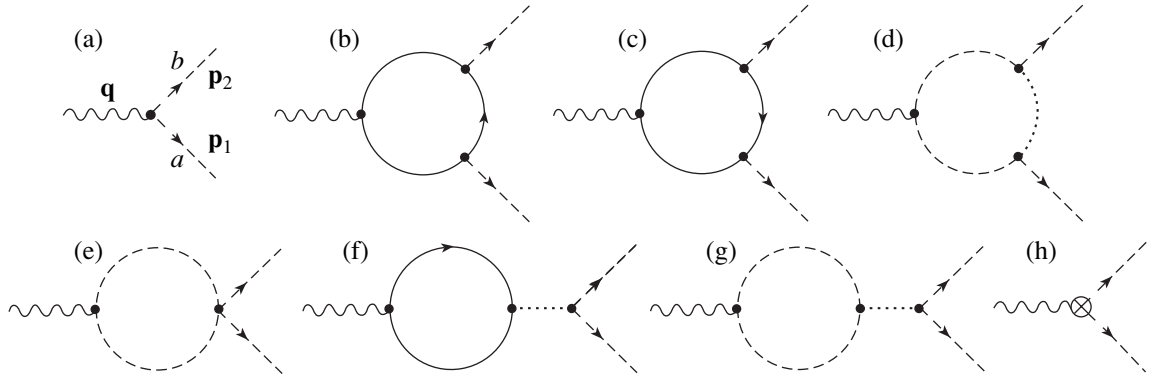


Fig. 1. One-loop diagrams contributing to the pion electromagnetic form factor in the σ-model. Dashed lines depict pion; dotted lines, sigma; solid lines, nucleon; and wavy lines, photon. Small crossed circle denotes the counterterm. Diagram (a) corresponds to the pion form factor in sQED.

Appendix A.4]). Exploiting conditions (10), we find the constant δ_{Z_π}:

$$\delta_{Z_\pi} = -\frac{g_\pi^2}{4\pi^2}\int_0^1\left[I_\epsilon - \ln\frac{\tilde\Delta_{N\pi}}{\Lambda^2} + \frac{m_\pi^2 x(1-x)}{\tilde\Delta_{N\pi}} + \frac{(m_\sigma^2-m_\pi^2)^2 x(1-x)}{4m_N^2\tilde\Delta_{\pi\sigma}}\right]dx, \tag{11}$$

$$\tilde\Delta_{N\pi} = m_N^2 - m_\pi^2 x(1-x),$$
$$\tilde\Delta_{\pi\sigma} = m_\sigma^2 x + m_\pi^2(1-x)^2. \tag{12}$$

In these equations,

$$I_\epsilon = \frac{2}{\epsilon} - \gamma_E + \ln 4\pi,$$

$$\epsilon = 4 - D \longrightarrow 0,$$

where D is the space–time dimension, $\gamma_E \approx 0.5772$ is the Euler constant, and Λ is an arbitrary scale mass, which drops out in the physical observables.

The one-loop contributions to the pion electromagnetic vertex coming from the σ-model are shown in Fig. 1. Using the isospin structure of the vertices, or negative charge-conjugation parity of the photon, one can show that diagrams (e), (f), and (g) vanish. Counterterm (h) cancels the divergences coming from loop contributions (b) and (c), while contribution (d) is finite.

In the general case of the off-mass-shell pions, the electromagnetic vertex $\Gamma^\mu_{ab}(p_1, p_2, q)$ for the process

$$\gamma^*(q) \longrightarrow \pi^a(p_1) + \pi^b(p_2)$$

has the form

$$-i\Gamma^\mu_{ab}(p_1, p_2, q) = \varepsilon^{3ab}[F(p_1^2, p_2^2, q^2)(p_2 - p_1)^\mu + G(p_1^2, p_2^2, q^2)(p_2 + p_1)^\mu], \tag{13}$$

with scalar functions $F(p_1^2, p_2^2, q^2)$ and $G(p_1^2, p_2^2, q^2)$. On the mass shell, $p_1^2 = p_2^2 = m_\pi^2$, the function $G(m_\pi^2, m_\pi^2, q^2)$ drops out, while $F(m_\pi^2, m_\pi^2, q^2)$ becomes equal to the pion form factor $F_\pi(q^2)$. With the loop corrections denoted by $\Delta F^{(\sigma)}(q^2)$, we find

$$F_\pi^{(\sigma)}(q^2) = 1 + \Delta F^{(\sigma)}(q^2) + \delta_{Z_\pi}, \tag{14}$$

where the total correction is finite and is given by

$$\Delta F^{(\sigma)}(q^2) + \delta_{Z_\pi} = \frac{g_\pi^2}{4\pi^2}\left[\int_0^1\left(\ln\frac{m_N^2 - m_\pi^2 x(1-x)}{m_N^2 - q^2 x(1-x)} - x(1-x)\left[\frac{m_\pi^2}{\tilde\Delta_{N\pi}} + \frac{(m_\sigma^2-m_\pi^2)^2}{4m_N^2\tilde\Delta_{\pi\sigma}}\right]\right)dx + \int_0^1\int_0^1\left(\frac{y^2 m_\pi^2}{\Delta_{N\pi}(y,x)} + \frac{y(1-y)(m_\sigma^2-m_\pi^2)^2}{4m_N^2\Delta_{\pi\sigma}(y,x)}\right)dxdy\right], \tag{15}$$

$$\Delta_{N\pi}(y,x) = m_N^2 - q^2 y^2 x(1-x) - m_\pi^2 y(1-y),$$
$$\Delta_{\pi\sigma}(y,x) = m_\sigma^2(1-y) - y^2(q^2 x(1-x) - m_\pi^2), \tag{16}$$

with $\tilde\Delta_{N\pi}$ and $\tilde\Delta_{\pi\sigma}$ defined in Eq. (12).

2.4. Contribution to the Pion Electromagnetic Form Factor from the ρ-Meson

The contribution to the pion electromagnetic form factor from the ρ-meson can be written in the compact form

$$F_\pi^{(\rho)}(q^2) = -\frac{g_\rho(q^2)}{f_\rho(q^2)}\frac{q^2}{q^2 - \hat{m}_\rho^2 - \Pi_p(q^2)} + \Delta F_\pi^{(\rho\omega)}(q^2). \quad (17)$$

This expression includes several effects coming from the loop corrections shown in Fig. 2.

(1) The q^2-dependent vertex $g_\rho(q^2)$ describes loop corrections to the $\rho\pi\pi$ coupling that originate from the σ-model (Fig. 2, diagrams (a)). These corrections have not been included in [11]. We can write

$$g_\rho(q^2) = g_\rho[1 + \Delta F^{(\sigma)}(q^2) + \delta_{Z_\pi}]. \quad (18)$$

It is seen that the expression in the square brackets is the same as in Eq. (14) and is finite. From Eq. (18), we obtain

$$g_\rho(q^2) = g_\rho(m_\rho^2)\frac{1 + \Delta F^{(\sigma)}(q^2) + \delta_{Z_\pi}}{1 + \Delta F^{(\sigma)}(m_\rho^2) + \delta_{Z_\pi}} \quad (19)$$

in terms of the constant $g_\rho(m_\rho^2)$. From the experimental width of the $\rho \longrightarrow \pi\pi$ decay, $\Gamma_{\rho\to\pi\pi} = 150.7$ MeV [15], we have $|g_\rho(m_\rho^2)| = 6.05$. To find the real and imaginary part of $g_\rho(m_\rho^2)$, we can use the relations

$$\mathrm{Re}\, g_\rho(m_\rho^2) = \frac{1}{\sqrt{1+\lambda^2}}\left|g_\rho(m_\rho^2)\right|,$$

$$\mathrm{Im}\, g_\rho(m_\rho^2) = \frac{\lambda}{\sqrt{1+\lambda^2}}\left|g_\rho(m_\rho^2)\right|, \quad (20)$$

$$\lambda = \mathrm{Im}\Delta F^{(\sigma)}(m_\rho^2)[1 + \mathrm{Re}\Delta F^{(\sigma)}(m_\rho^2) + \delta_{Z_\pi}]^{-1}.$$

(2) The ρ-meson self-energy has the structure

$$\Pi_\rho^{\mu\nu}(q) = (g^{\mu\nu} - q^\mu q^\nu/q^2)\Pi_\rho(q^2)$$

and corresponds to the diagrams shown in Fig. 2b. It leads to the following exact propagator of the ρ-meson (Fig. 2, diagrams (c)):

$$G_\rho^{\mu\nu}(q) = -i\frac{g^{\mu\nu} - q^\mu q^\nu/q^2}{q^2 - \hat{m}_\rho^2 - \Pi_\rho(q^2)} + i\frac{q^\mu q^\nu}{q^2\hat{m}_\rho^2}. \quad (21)$$

Calculation of the loop integrals in Fig. 2b results in

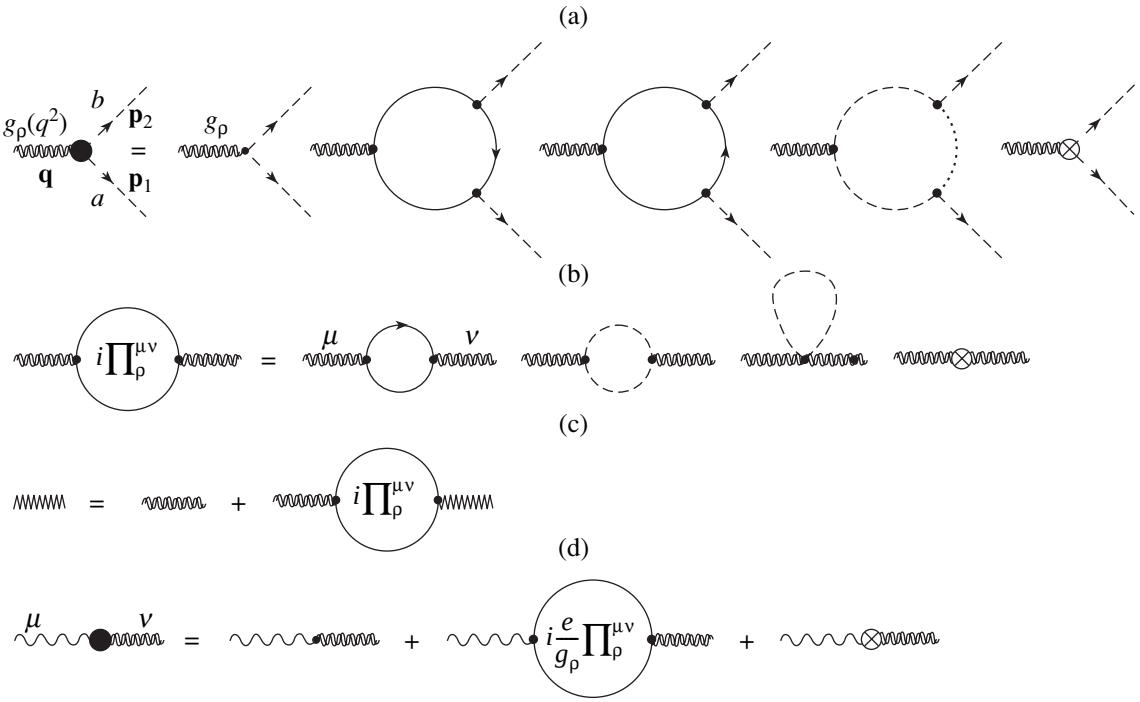


Fig. 2. Diagrams (a) show π, σ, and *N* loop corrections to the ρππ vertex; diagrams (b) show the π and *N* loop corrections to the self-energy of the ρ-meson; diagrams (c) graphically represent the equation for the exact ρ-meson propagator; diagrams (d) represent the π and *N* loop corrections to the γρ vertex. Double wavy lines depict the ρ-meson. The corresponding counterterms are indicated by crossed circles.

$$\Pi_\rho(q^2) = q^2[\omega(q^2) - \delta_{Z_\rho}],$$
$$\omega(q^2) = \frac{g_\rho^2}{16\pi^2}[-I_\epsilon + I(q^2)], \quad (22)$$

$$I(q^2) = 2\int_0^1 \left[2x\ln\frac{\Delta N}{\Lambda^2} + (1-2x)\ln\frac{\Delta_\pi}{\Lambda^2}\right](1-x)dx, \quad (23)$$

with

$$\Delta_N = m_N^2 - q^2x(1-x), \quad \Delta_\pi = m_\pi^2 - q^2x(1-x).$$

The self-energy has a logarithmic divergence and requires renormalization. The authors of [11] renormalized the self-energy by applying a dispersion relation with two subtractions. We prefer an alternative method of counterterms, which is expressed in Eq. (22) via the constant δ_{Z_ρ}. We can fix the latter from the constraint on the self-energy at physical mass m_ρ of the ρ-meson,

$$\frac{d}{dq^2}\mathrm{Re}\Pi_\rho(q^2)\Big|_{q^2 = m_\rho^2} = 0 \quad (24)$$

$$\Rightarrow \delta_{Z_\rho} = Z_\rho - 1 = \mathrm{Re}\omega(m_\rho^2) + m_\rho^2\mathrm{Re}\omega'(m_\rho^2), \quad (25)$$

where

$$\mathrm{Re}\omega'(q^2) \equiv \frac{d}{dq^2}\mathrm{Re}\omega(q^2).$$

It is seen from Eqs. (22) and (25) that the self-energy

$$\Pi_\rho(q^2) = q^2[\omega(q^2) - \mathrm{Re}\omega(m_\rho^2) - m_\rho^2\mathrm{Re}\omega'(m_\rho^2)]$$

is finite. Near the physical mass, it has the expansion

$$\mathrm{Re}\Pi_\rho(q^2) = -m_\rho^4\mathrm{Re}\omega'(m_\rho^2) + O((q^2 - m_\rho^2)^2), \quad (26)$$

and therefore the coupling g_ρ is not renormalized due to self-energy loops [11]. There is also a finite mass shift:

$$m_\rho^2 - \hat{m}_\rho^2 = -m_\rho^4\mathrm{Re}\omega'(m_\rho^2). \quad (27)$$

For the definition of $\hat{m}_\rho$, see the paragraph before Eq. (8).

Above the two-pion threshold, the self-energy acquires an imaginary part that comes from the pion loop (the third diagram in Fig. 2b). Namely, at $q^2 < 4m_N^2$, the imaginary part and the q^2-dependent $\rho \longrightarrow \pi\pi$ decay width are given by the respective expressions

$$\mathrm{Im}\Pi(q^2) = -\frac{g_\rho^2 q^2}{48\pi}\left(1 - \frac{4m_\pi^2}{q^2}\right)^{3/2}\theta(q^2 - 4m_\pi^2),$$
$$\Gamma_\rho(q^2) = -\mathrm{Im}\Pi(q^2)/\sqrt{q^2}. \quad (28)$$

(3) Closely related to the self-energy are loop corrections to the γρ coupling constant shown in Fig. 2d. The sum of all contributions is proportional to the tensor $g^{\mu\nu} - q^\mu q^\nu/q^2$, similarly to the tree-level term. Introducing the q^2-dependent vertex, we obtain

$$\frac{1}{f_\rho(q^2)} = \frac{1}{f_\rho} - \frac{\omega(q^2)}{g_\rho} + \delta_{f_\rho}, \quad (29)$$

where δ_{f_ρ} can be fixed by requiring that on the mass shell $q^2 = m_\rho^2$, the coupling $f_\rho(m_\rho^2)$ is related to the $\rho \longrightarrow e^+e^-$ decay width. The experimental width $\Gamma_{\rho\to e^+e^-} = 6.77$ keV [15] is reproduced with $|f_\rho(m_\rho^2)| \approx 5.03$. From Eq. (29), we find

$$\frac{1}{f_\rho(q^2)} = \frac{1}{f_\rho(m_\rho^2)} + \frac{1}{g_\rho}[\omega(m_\rho^2) - \omega(q^2)], \quad (30)$$

where the real part of the constant $f_\rho(m_\rho^2)$ is determined from $|f_\rho(m_\rho^2)|$ and from the imaginary part,

$$\mathrm{Im}\, f_\rho(m_\rho^2) = \left|f_\rho(m_\rho^2)\right|^2\mathrm{Im}\omega(m_\rho^2)/g_\rho.$$

It is seen from (30) that the effective γρ vertex is finite. A similar procedure for this vertex was used in [11], although only the real part of $f_\rho(m_\rho^2)$ was taken from experiment.

We also mention that, in calculating $\Pi_\rho(q^2)$ and $f_\rho(q^2)$, we used $|g_\rho(m_\rho^2)|$ instead of g_ρ in order to obtain the correct width of the ρ-meson.

(4) The last term in Eq. (17) describes the ρ–ω interference due to electromagnetic effects [8]. The explicit form of the contribution to the pion form factor can be taken from [16],

$$\Delta F_\pi^{(\rho\omega)}(q^2) = -\varepsilon_{\rho\omega}\frac{g_\rho}{f_\omega}\frac{q^2}{q^2 - m_\omega^2 + im_\omega\Gamma_\omega}, \quad (31)$$

$$\varepsilon_{\rho\omega} = \frac{\Pi_{\rho\omega}}{m_\omega^2 - m_\rho^2 - i[m_\omega\Gamma_\omega - m_\rho\Gamma_\rho(q^2)]}, \quad (32)$$

where $\Gamma_\omega = 8.43$ MeV is the full decay width of the ω-meson with mass m_ω, $f_\omega = 17.05$ is the γω coupling constant, which is fixed from the $\omega \longrightarrow e^+e^-$ decay width $\Gamma_{\omega\to e^+e^-} = 0.6$ keV [15], and $\Pi_{\rho\omega} \approx -3.8 \times 10^{-3}$ GeV^2 is the mixed ρ–ω self-energy.

3. RESULTS AND DISCUSSION

We first specify the parameters of the model. The constant g_π is determined from the tree-level Gold-

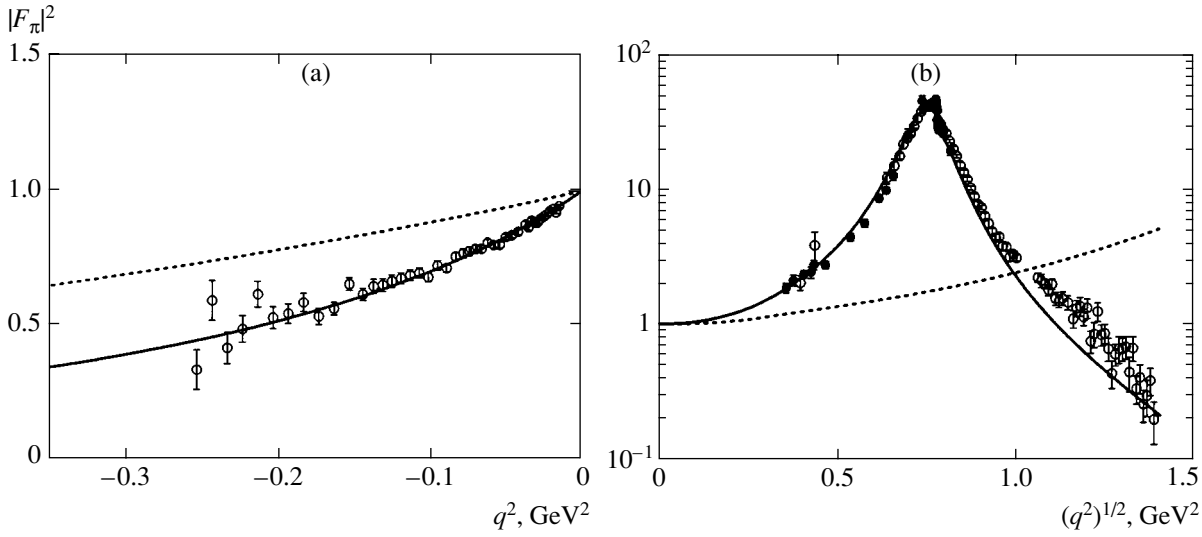


Fig. 3. The pion electromagnetic form factor for spacelike q^2 (a) and timelike q^2 (b). Experimental data are from [18] and [19], respectively. Solid lines, σ-model + VMD; dotted lines, σ-model.

berger–Treiman relation

$$g_\pi = m_N/f_\pi$$

(the first equation in (4)), while $g_\rho(m_\rho^2)$ and $f_\rho(m_\rho^2)$ are fixed from experiment, as described in Section 2.4. The σ mass is chosen equal to the mass of ρ, i.e., $m_\sigma = m_\rho$, in line with [17], where σ and ρ are assumed degenerate. Furthermore, $m_\rho = 768.5$ MeV and $m_\omega = 782.57$ MeV [15].

It is interesting to note that calculation of the self-energy of the ρ-meson gives $\hat{m}_\rho = 795$ MeV. This value is rather close to the physical mass m_ρ. In this connection, we note that the authors of [11] fitted $\hat{m}_\rho$ from the ππ scattering and obtained 810 MeV. The difference in the above values of $\hat{m}_\rho$ is partially due to our taking the nucleon loop into account, which was not considered in [11].

As mentioned in Section 2.4, both the real and imaginary parts of the coupling constants $g_\rho(m_\rho^2)$ and $f_\rho(m_\rho^2)$ have been included. The calculation yields

$$\mathrm{Re}\, g_\rho(m_\rho^2) = 6.036, \quad \mathrm{Im}\, g_\rho(m_\rho^2) = 0.405,$$

$$\mathrm{Re}\, f_\rho(m_\rho^2) = 4.96, \quad \mathrm{Im}\, f_\rho(m_\rho^2) = -0.82.$$

Taking the imaginary parts into account leads to a small correction to the results obtained with

$$\mathrm{Im}\, g_\rho(m_\rho^2) = \mathrm{Im}\, f_\rho(m_\rho^2) = 0.$$

Our main results are demonstrated in Fig. 3 and Tables 1 and 2. The calculated pion form factor $|F_\pi(q^2)|^2$ for spacelike and timelike values of q^2 is presented in Fig. 3. Apparently, the agreement with the data [18] from elastic electron–pion scattering and the data [19] from e^-e^+ annihilation in two pions is quite good. We emphasize that there are no fitting or tuning parameters in our approach.

There is a strong interference of the two contributions, $F_\pi^{(\sigma)}$ and $F_\pi^{(\rho)}$, in the total form factor. In Fig. 3, we show the contribution $F_\pi^{(\sigma)}$ separately. It follows from Eqs. (17) and (18) that the ρ contribution also includes π, σ, and N loops coming from the σ-model. Switching off these corrections, i.e., setting

$$\Delta F^{(\sigma)}(q^2) + \delta_{Z_\pi} = 0,$$

we obtain

$$F_\pi^{(sQED+VMD)}(q^2) = 1 - \frac{g_\rho}{f_\rho(q^2)}\frac{q^2}{q^2 - \hat{m}_\rho^2 - \Pi_\rho(q^2)} + \Delta F_\pi^{(\rho\omega)}(q^2). \quad (33)$$

Here, the first term corresponds to sQED, and the sec-

Table 1. Two-pion contribution $a_\mu^{(\mathrm{had},\pi)}$ to the muon anomalous magnetic moment (in units 10^{-11}). The upper integration limit in Eq. (2) is 0.8 GeV^2

sQED	σ-model	sQED + VMD	σ-model + VMD	σ-model + VMD ($f_\rho = g_\rho$)	Ref. [2]
525	753	4667	4763	4745	4774 ± 51

Table 2. Dependence of $a_\mu^{(\mathrm{had},\pi)}$ on the mass of σ-meson

m_σ, GeV	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2
$a_\mu^{(\mathrm{had},\pi)}$, 10^{-11}	4546	4583	4640	4710	4788	4867	4946	5024	5099

ond and the third terms are the ρ-meson contributions. We note that Eq. (33) corresponds to the extended version of VMD in [11]; in the "standard" VMD model [8], we have the dependence

$$\frac{m_\rho^2}{m_\rho^2 - q^2 - im_\rho\Gamma_\rho(q^2)}.$$

Our calculation shows that the difference between the form factor calculated in σ-model + VMD (Eqs. (14) and (17)), and that in sQED + VMD (Eq. (33)), is small, and therefore the results for sQED + VMD are not plotted in Fig. 3. Nevertheless, the difference may show up in the integrated quantity $a_\mu^{(had,\pi)}$ for the muon AMM.

The calculated values of $a_\mu^{(had,\pi)}$ are shown in Table 1. In general, loop corrections are important in the σ-model (compare the first and the second columns). Their role is, however, diminished in the full calculation, which includes the dominant ρ-meson contribution (the third and the fourth columns in Table 1). The difference between σ-model + VMD and sQED + VMD calculations is about 2%.

In this connection, our result can be used to estimate the size of radiative corrections due to final-state-photon radiation in the $e^-e^+ \longrightarrow \pi^+\pi^-\gamma$ process. The corresponding contribution to the muon AMM, $a_\mu^{(had,\pi\gamma)}$, is calculated in [4] in the sQED + VMD framework. In general, the ratio

$$\sigma(e^+e^- \longrightarrow \pi^+\pi^-\gamma)/\sigma(e^+e^- \longrightarrow \pi^+\pi^-)$$

is on the order of α. We expect that the model dependence of $a_\mu^{(had,\pi\gamma)}$ is similar to that of $a_\mu^{(had,\pi)}$. Therefore, the deviation of $a_\mu^{(had,\pi\gamma)}$ calculated in sQED + VMD from $a_\mu^{(had,\pi\gamma)}$ in a more realistic model, for example, in σ-model + VMD, is about 2%. The overall model-dependent effect in the contribution $a_\mu^{(had,\pi\gamma)}$ to the muon AMM is of the order $\alpha \times 2\% \approx 0.015\%$ and is therefore negligible.

In the fifth column, we show the result obtained if we set $f_\rho = g_\rho$ in Lagrangian (6). This approximation corresponds to the full universality of Sakurai. In this case, the renormalization procedure for the γρ vertex changes and δ_{f_ρ} in Eq. (8) is equal to δ_{Z_ρ}/g_ρ. The numerical results for the form factor and $a_\mu^{(had,\pi)}$, however, change very little, e.g., by about 0.4% for the integral.

As mentioned above, we chose the mass $m_\sigma = m_\rho$ for the σ-meson in the calculation. This particle is associated with the $f_0(400–1200)$-meson in [15]. In view of its undetermined status, we study the dependence of the calculated integral $a_\mu^{(had,\pi)}$ on m_σ in Table 2. As can be seen from Table 2, the integral varies considerably. We take the value 4774 ± 51 from [2] as a very accurate fit to the experimental integral. Then, for the indicated error bars, we obtain the mass of σ in the interval from 720 to 850 MeV. The central value 785 MeV is surprisingly close to the ρ-meson mass. Therefore, our calculation agrees with the hypothesis in [17] about degeneracy of σ- and ρ-mesons.

In the calculations, we took only the diagrams with the ρ-meson entering on the tree level into account. In particular, the loops with an intermediate ρ-meson for the $\gamma\pi^+\pi^-$ and $\rho\pi^+\pi^-$ vertices are left out. Such contributions can be consistently considered in the models in which the ρ-meson is included together with its chiral partner, the axial vector a_1-meson, for example, in the so-called gauged σ-model [20] or chiral quantum hadrodynamics [21, 22]. This work requires further investigation.

4. CONCLUSIONS

We developed a model for the electromagnetic vertex of the pion. The model is based on the linear σ-model, which generates loops with the intermediate pion, sigma, and nucleon. The ρ-meson is included in line with the extended VMD model [11]. The coupling of ρ to the pion and nucleon is introduced via gauge-covariant derivatives, and the direct γρ coupling has a gauge-invariant form. The ρ-meson self-energy and the modified γρ vertex are generated by the pion and nucleon loops. The renormalization is consistently performed using the method of counterterms without cutoff parameters.

The pion electromagnetic form factor calculated in the one-loop approximation in strong interaction agrees well with the precise data obtained from elastic $e^-\pi^+$ scattering and e^+e^- annihilation into $\pi^+\pi^-$. The effect of the σ-model loops turns out to be small.

We calculated the contribution of the $e^+e^- \longrightarrow \pi^+\pi^-$ process to the muon AMM, $a_\mu^{(had,\pi)}$. The calculation agrees quite well (by 0.15%) with the recent, very accu-

rate fit in [2]. The contribution of the σ-model loops to $a_\mu^{(had,\pi)}$ is about 2%.

We also estimated the size of the model-dependent effects in $a_\mu^{(had,\pi\gamma)}$, the contribution to the muon AMM from final-state-photon radiation in the $e^+e^- \longrightarrow \pi^+\pi^-\gamma$ process. It is about $\alpha \times 2\% \approx 0.015\%$ and is therefore negligible. Hence, our calculation does not contradict the conclusion in [4] that the final-state radiative process $e^+e^- \longrightarrow \pi^+\pi^-\gamma$ can be evaluated in scalar QED supplemented with the VMD model.

The only free parameter of the model, which is not fixed from the experiment, is the σ-meson mass. Comparison with the fit in [2] strongly indicates that the value of this mass is close to the mass of the ρ-meson. This conclusion is consistent with [17], where the mesons σ and ρ are assumed to be degenerate.

REFERENCES

1. H. N. Brown *et al.*, Phys. Rev. Lett. **86**, 2227 (2001); G. W. Bennett, B. Bousquet, H. N. Brown, *et al.*, Phys. Rev. Lett. **89**, 101804 (2002).
2. J. F. de Trocóniz and F. J. Ynduráin, Phys. Rev. D **65**, 093001 (2002); hep-ph/0106025.
3. S. J. Brodsky and E. de Rafael, Phys. Rev. **168**, 1620 (1968).
4. H. Czyz, A. Grezelinska, J. H. Kühn, and G. Rodrigo, hep-ph/0308312; Eur. Phys. J. (in press).
5. A. B. Arbuzov, E. A. Kuraev, N. P. Merenkov, and L. Trentadue, J. High Energy Phys. **12**, 009 (1998); hep-ph/9804430; S. Binner, J. H. Kühn, and K. Melnikov, Phys. Lett. B **459**, 279 (1999); hep-ph/9902399.
6. A. Hoefer, S. Jadach, J. Gluza, and F. Jegerlehner, hep-ph/0302258.
7. J. S. Schwinger, *Particles, Sources, and Fields* (Addison-Wesley, Redwood City, 1989), Vol. 3, p. 99; M. Drees and K. I. Hikasa, Phys. Lett. B **252**, 127 (1990).
8. R. P. Feynman, *Photon–Hadron Interactions* (Benjamin, Reading, Mass., 1972), Chap. 6.
9. M. Gell-Mann and M. Lévy, Nuovo Cimento **16**, 705 (1960).
10. M. Herrmann, B. L. Friman, and W. Nörenberg, Nucl. Phys. A **560**, 411 (1993).
11. E. Klingle, N. Kaizer, and W. Weise, Z. Phys. A **356**, 193 (1996).
12. V. De Alfaro, S. Fubini, G. Furlan, and S. Rossetti, *Currents in Hadron Physics* (North-Holland, Amsterdam, 1973).
13. E. M. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory* (Addison-Wesley, Reading, Mass., 1995).
14. C. Itzykson and J. B. Zuber, *Quantum Field Theory* (McGraw-Hill, New York, 1980; Mir, Moscow, 1984).
15. K. Hagiwara *et al.* (Particle Data Group), Phys. Rev. D **66**, 010001 (2002); URL:http://pdg.lbl.gov.
16. H. B. O'Connell, B. C. Pearce, A. W. Thomas, and A. G. Williams, Phys. Lett. B **354**, 14 (1995).
17. S. Weinberg, Phys. Rev. Lett. **65**, 1177 (1990).
18. S. R. Amendolia *et al.*, Nucl. Phys. B **277**, 168 (1986).
19. L. M. Barkov *et al.*, Nucl. Phys. B **256**, 365 (1985).
20. B. W. Lee and H. T. Nieh, Phys. Rev. **166**, 1507 (1968).
21. B. D. Serot and J. D. Walecka, Acta Phys. Pol. B **21**, 655 (1992).
22. A. Yu. Korchin, D. Van Neck, and M. Waroquier, Phys. Rev. C **67**, 015207 (2003); nucl-th/0302042.

Journal of Experimental and Theoretical Physics, Vol. 99, No. 2, 2004, pp. 234–253.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 2, 2004, pp. 268–287.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Higher Order Corrections to the Lipatov Asymptotics in the φ^4 Theory

D. A. Lobaskin and I. M. Suslov*
Kapitza Institute of Physical Problems, Russian Academy of Sciences, Moscow, 117334 Russia
**e-mail: suslov@kapitza.ras.ru*
Received February 11, 2004

Abstract—Higher orders in perturbation theory can be calculated by the Lipatov method [1]. For most field theories, the Lipatov asymptotics has the functional form $ca^N\Gamma(N + b)$ (N is the perturbation theory order); relative corrections to this asymptotics have the form of a power series in $1/N$. The coefficients of higher order terms of this series can be calculated using a procedure analogous to the Lipatov approach and are determined by the second instanton in the field theory in question. These coefficients are calculated quantitatively for the n-component φ^4 theory under the assumption that the second instanton is (i) a combination of elementary instantons and (ii) a spherically asymmetric localized function. A technique of two-instanton computations, as well as the method for integrating over rotations of an asymmetric instanton in the coordinate state, is developed. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION AND MAIN RESULTS

According to Lipatov [1], higher orders of perturbation theory are determined by saddle-point configurations (instantons) of the corresponding functional integrals. A typical asymptotics for coefficients Z_N in the expansion of a certain quantity $Z(g)$ in the coupling constant g,

$$Z(g) = \sum_{N=0}^{\infty} Z_N g^N, \tag{1}$$

has the form

$$Z_N = cS_0^{-N}\Gamma(N + b), \quad N \longrightarrow \infty, \tag{2}$$

where S_0 is the instanton action and b and c are constants. The corrections to the Lipatov asymptotics (2) have the form of a regular expansion in $1/N$,

$$Z_N = cS_0^{-N}\Gamma(N + b) \times\left\{1 + \frac{A_1}{N} + \frac{A_2}{N^2} + \ldots + \frac{A_K}{N^K} + \ldots\right\}; \tag{3}$$

the calculation of these corrections provides additional information on the coefficient function Z_N. It was proved recently by one of the authors [2] that the series (3) diverges factorially and the typical asymptotics of coefficients A_K for $K \longrightarrow \infty$ has the form

$$A_K = \tilde{c}\left(\ln\frac{S_1}{S_0}\right)^{-K}\Gamma\left(K + \frac{r' - r}{2}\right), \tag{4}$$

where S_0 and S_1 are the values of the first and second instantons in the given field theory, while r and r' are the corresponding numbers of zeroth modes; we assume that instantons are labelled in the order of increasing the corresponding action. This study is aimed at obtaining close results for the asymptotic form A_K in the n-component φ^4 theory with the action

$$S\{g, \varphi\} = \int d^d x\left\{\frac{1}{2}\sum_{\alpha=1}^{n}[\nabla\varphi_\alpha(x)]^2 + \frac{1}{2}m^2\sum_{\alpha=1}^{n}\varphi_\alpha^2(x) + \frac{1}{4}g\left(\sum_{\alpha=1}^{n}\varphi_\alpha^2(x)\right)^2\right\}, \tag{5}$$

where d is the dimension of the space. Initially, we planned to calculate the constant in formula (4); however, it was found in the course of analysis that the functional form of this constant in the presence of soft modes requires correction.

Unfortunately, complete information on higher order instantons in the φ^4 theory is not available. It is known [3] that the zero-node spherically symmetric instanton derived analytically for $d = 1, 4$ and numerically for $d = 2, 3$ [4] possesses the minimal action S_0. Configurations with actions $2S_0$, $3S_0$, etc., corresponding to several infinitely remote elementary instantons, obviously also exist; for $d = 1$, such configurations cover the entire instanton spectrum. Higher order spherically symmetric instantons are higher lying[1] for

[1] The second spherically symmetric instanton has an action $6.6S_0$ for $d = 2$ and $6.3S_0$ for $d = 3$.

$d = 2, 3$ and are missing for $d = 4$. A group of scientists headed by Eleonsky [5] attempted to find instantons that do not possess spherical symmetry; two instantons with an action higher than that for the third spherically symmetric instanton were found for $d = 2$, while no nontrivial instantons were discovered for $d = 3$. Ushveridze [6] found analytically a series of asymmetric instantons for $d = 4$, which begins with[2] $8S_0$. A numerical algorithm for determining the "main sequence" of instantons was proposed in [7]. The realization of this algorithm shows[3] that lower order instantons in this sequence split into elementary instantons. Thus, the most probable candidate for the role of the second instanton in the φ^4 theory is a combination of two elementary instantons; the present study is mainly based on this assumption. However, since the existence of an asymmetric instanton with an action smaller than $2S_0$ cannot be ruled out, formal results corresponding to this case are given in Section 8.

It will be proved below that result (4) is valid when the "equipartition law" is applicable, i.e., when all fluctuation modes can be completely separated into the zeroth and vibrational modes (Section 2). For two-instanton configurations, a soft mode corresponding to a change in the distance between elementary instantons and reducible to vibrations at a nonanalytic minimum inevitably exists. As a result, the right-hand side of formula (4) acquires logarithmic corrections for $d = 1$, 2, and 3 and even power corrections for $d = 4$.

Calculations for two-instanton configurations by the saddle-point method were considered in [8–12] in connection with the Lipatov asymptotics for problems with degenerate vacuum (such as quantum chromodynamics problems). In this case, the main difficulty lies in the emergence of poorly defined integrals, which were interpreted in [8–12] at the level of heuristic receptors, the interpretation being not quite consistent even according to the authors of these publications. In the computation of the asymptotic form of A_K, this problem acquires a new aspect and requires a thorough analysis. For this reason, we begin with the discussion of the Bogomolny–Parisi dispersion relation [13, 14], which is a source of poorly defined expressions (Section 3). On the basis of this relation, the general interrelation between the corrections to the asymptotics and higher order instantons, which was heuristically pointed out in [2], is established (Section 4). Then the combination rule for instantons is derived (Section 5) and the general computational algorithm in the presence of soft modes is formulated (Section 6) and subsequently applied to the φ^4 theory (Section 7). The results for an asymmetric second instanton are considered in Section 8.

[2] The value $27/16S_0$ obtained in [2] is erroneous.

[3] E.R. Podolyak, private communication.

We will study functional integrals of the form

$$Z_M(g) = \int D\varphi\varphi_{\alpha_1}(x_1)\varphi_{\alpha_2}(x_2)\ \ldots\ \varphi_{\alpha_M}(x_M)\exp(-S\{g, \varphi\}), \tag{6}$$

via which M-point Green functions can be expressed,

$$G_M(g) = \frac{Z_M(g)}{Z_0(g)}. \tag{7}$$

We have derived the following expression for the asymptotic form of coefficients A_K corresponding to the Green function $G_M(g)$ for $d = 1$:

$$A_K = -\frac{2^{-M/2}}{(\pi/2)\Gamma(n/2)}\left(\frac{3}{2\ln 2}\right)^{n/2} \times \Gamma\left(K + \frac{n}{2}\right)(\ln 2)^{-K}[\ln K + C], \tag{8}$$

$$C = C_E + \ln\left(\frac{6}{\ln 2}\right) + \frac{\psi(1/2) - \psi(n/2)}{2};$$

here, C_E is the Euler constant and $\psi(x)$ is the logarithmic derivative of the gamma function. For $d = 2$, to within logarithmic accuracy, we have

$$A_K = -\frac{2^{-M/2}}{19.7}\frac{(0.702)^n}{\Gamma(n/2)} \times \Gamma\left(K + \frac{n+1}{2}\right)(\ln 2)^{-K}\ln^2 K; \tag{9}$$

similarly, for $d = 3$, we have

$$A_K = -\frac{2^{-M/2}}{2.12}\frac{(0.704)^n}{\Gamma(n/2)} \times \Gamma\left(K + \frac{n+2}{2}\right)(\ln 2)^{-K}\ln^3 K. \tag{10}$$

For $d = 4$, the results depend on the coordinates appearing in the Green function and are rather cumbersome (see Section 7). The results are slightly simplified as we pass to the momentum representation and choose momenta p_i corresponding to a symmetric point ($p_i \sim p$):

$$A_K = B\exp\left(\nu\ln\frac{\mu}{p}\right) \times \Gamma\left(K + \frac{n+4}{2} + \nu\right)(\ln 2)^{-K}, \tag{11}$$

where μ is the point of charge normalization, $\nu = (n + 8)/3$, and the values of constant B are given in the table. In the scalar case ($n = 1$), the main contribution to the

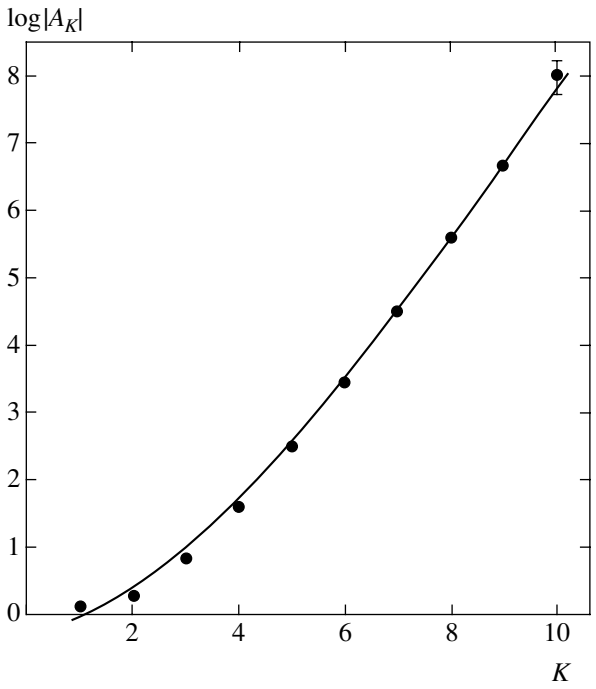


Fig. 1. Comparison of asymptotic formula (12) (solid curve) with coefficients A_K determined numerically in [15] (circles)

asymptotic form vanishes and we should expect the behavior corresponding to the next order in $1/K$:

$$A_K = \text{const} \cdot \exp\left(\nu \ln\frac{\mu}{p}\right) \times \Gamma\left(K + \frac{n+4}{2} + \nu - 1\right)(\ln 2)^{-K}. \quad (11')$$

The results for the logarithm of the vacuum integral $Z_0(g)$ can formally be obtained from expressions (8)–(11) by substituting $M = 0$ and introducing an additional factor 1/2 into the right-hand sides of these expressions. In particular, for the ground-state energy of a harmonic oscillator ($d = 1$, $n = 1$), we obtain

$$A_K = -\frac{\ln K + 2.74}{3.78}\Gamma\left(K + \frac{1}{2}\right)(\ln 2)^{-K}, \quad (12)$$

which can be compared with the results obtained by Bender and Wu [15] (Fig. 1); in contrast to [2], this comparison is carried out without any fitting parameters.

Values of parameter B in formula (11)

n	$B \times 10^4$	
	$M = 2$	$M = 4$
0	–9.05	–8.72
1	0	0
2	3.25	1.45
3	4.55	1.50

For an asymmetric instanton (Section 8), soft modes are absent and the structure of the result can be predicted beforehand proceeding from formula (4). For the first instanton of the φ^4 theory, zero modes contain d translational, $n - 1$ rotational (associated with a change in the direction of the vector field φ) and (for $d = 4$) one dilatational mode corresponding to a change in the instanton radius.[4] For the second instanton, in view of its low symmetry, $d(d-1)/2$ additional modes associated with rotations in the coordinate space appear; consequently, we have

$$A_K = \tilde{c}\left[\ln\left(\frac{S_1}{S_0}\right)\right]^{-K}\Gamma\left(K + \frac{d(d-1)}{4}\right). \quad (13)$$

The latter modes have not been considered before, which makes the calculation of constant $\tilde{c}$ (Section 8) nontrivial from the methodical point of view. The method for integrating with respect to these modes is of interest for quantum electrodynamics, where even the first instanton is asymmetric [16].

2. STRUCTURE OF THE SADDLE-POINT CONTRIBUTION: EQUIPARTITION LAW

In the subsequent analysis, we will use the brief notation for integral (6),

$$Z(g) = \int D\varphi\varphi^{(1)}\ldots\varphi^{(M)}\exp(-S\{g,\varphi\}), \quad (14)$$

and will normalize this integral to an analogous integral with $M = 0$, $g = 0$, including factor $Z_0^{-1}(0)$ in the symbol $D\varphi$. Using the properties of action uniformity,[5] which are typical of the φ^4 theory,

$$S\{g,\varphi\} \longrightarrow \frac{S\{\varphi\}}{g} \text{ for } \varphi \longrightarrow \frac{\phi}{\sqrt{g}}, \quad (15)$$

and introducing the saddle-point configuration via the condition $S'\{\phi_c\} = 0$, in the vicinity of the saddle-

[4] For a two-instanton configuration, the number of zeroth modes is doubled ($r' = 2r$); by virtue of Eq. (4), this makes the contribution $r/2$ to the argument of the gamma-function, which is equal to $(n - 1 + d)/2$ for $d < 4$ (formulas (8)–(10)) and $(n + 4)/2$ for $d = 4$ (formula (11)).

[5] Analogous properties of uniformity are also used in other field theories; consequently, the subsequent analysis is also applicable to these theories after slight modifications.

point we have

$$Z(g) = g^{-M/2}\int D\varphi\phi_c^{(1)}\ldots\phi_c^{(M)} \times \exp\left(-\frac{S\{\phi_c\}}{g} - \frac{1}{2}(\delta\varphi, S''\{\phi_c\}\delta\varphi)\right) \quad (16)$$

($\delta\varphi = \varphi - \varphi_c$, $\varphi_c = \phi_c g^{-1/2}$); in the absence of zero modes, this leads to

$$Z(g) = \text{const}\cdot g^{-M/2}\exp\left(-\frac{S\{\phi_c\}}{g}\right). \quad (17)$$

We use the symbolic notation introduced in [2], where the prime and double prime denote the first and second functional derivatives, which are treated as a vector and a linear operator, respectively, while variables φ_i included in the symbol $D\varphi$ are assumed to be components of vector φ.

Expansion coefficients Z_N are defined by the integral

$$Z_N = \int_C \frac{dg}{2\pi i}\frac{Z(g)}{g^{N+1}} = \int_C \frac{dg}{2\pi i g}\int D\varphi\varphi^{(1)}\ldots\varphi^{(M)}\exp\left(-\frac{S\{\phi\}}{g} - N\ln g\right), \quad (18)$$

where contour C circumvents the point $g = 0$ in the positive direction. According to Lipatov [1], for large values of N, the integral can be evaluated by the saddle-point method. Introducing the saddle-point configuration via the conditions

$$S'\{\phi_c\} = 0, \quad g_c = \frac{S\{\phi_c\}}{N} \quad (19)$$

and carrying out the expansion in the vicinity of this point, we obtain

$$Z_N = e^{-N}g_c^{-N-M/2}\int_{-\infty}^{\infty}\frac{dt}{2\pi}\int D\varphi\phi_c^{(1)}\ldots\phi_c^{(M)} \times \exp\left(\frac{Nt^2}{2} - \frac{1}{2}(\delta\varphi, S''\{\phi_c\}\delta\varphi)\right), \quad (20)$$

where $g = g_c + ig_c t$; in the absence of zero modes, this gives

$$Z_N = \text{const}\cdot S\{\phi_c\}^{-N}\Gamma\left(N + \frac{M}{2}\right).$$

In fact, the functional integral always contains zero modes; for correct integration with respect to these modes, we introduce collective variables λ_i (such as the center of an instanton, its orientation, etc.), which are formally determined for an arbitrary configuration of φ and are its functionals, $\lambda_i = f_i\{\varphi\}$. The latter can be treated as uniform in φ with zero degree of uniformity [2]. We introduce into the integrand of integral (14) the expansion of unity,

$$1 = \prod_{i=1}^{r}\int d\lambda_i\delta(\lambda_i - f_i\{\varphi\}) = \prod_{i=1}^{r}\int d\lambda_i\delta(\lambda_i - f_i\{\varphi_c\} - (f'\{\varphi_c\}, \delta\varphi)) \quad (21)$$
$$= \prod_{i=1}^{r}\int d\lambda_i\delta(\lambda_i - f_i\{\phi_c\} - \sqrt{g}\{f'\{\phi_c\}, \delta\varphi\}),$$

where r is the number of zero modes. Using the degrees of freedom corresponding to zero modes, we choose the instanton from the condition $\lambda_i = f_i\{\phi_c\}$, after which ϕ_c becomes a function of λ_i (i.e., $\phi_c \equiv \phi_\lambda$). Simple calculations lead to the following results (see [2] for details):

$$Z(g) = c_g g^{-(M+r)/2}\exp\left(-\frac{S\{\phi_c\}}{g}\right), \quad (22)$$

$$c_g = \sqrt{\frac{\det S''\{0\}}{\det[S''\{\phi_c\}]_{P'}}}\frac{(2\pi)^{-r/2}}{\det[f'\{\phi_c\}]_P} \times \int\prod_{i=1}^{r}d\lambda_i\phi_\lambda^{(1)}\ldots\phi_\lambda^{(M)}, \quad (23)$$

$$Z_N = cS\{\phi_c\}^{-N}\Gamma\left(N + \frac{M+r}{2}\right), \quad (24)$$

$$c = \frac{S\{\phi_c\}^{-(M+r)/2}}{(2\pi)^{1+r/2}}\sqrt{-\frac{\det S''\{0\}}{\det[S''\{\phi_c\}]_{P'}}} \times \frac{1}{\det[f'\{\phi_c\}]_P}\int\prod_{i=1}^{r}d\lambda_i\phi_\lambda^{(1)}\ldots\phi_\lambda^{(M)}, \quad (25)$$

where $f'\{\phi_c\}$ is the operator whose matrix consists of columns $f_i'\{\phi_c\}$ and subscripts P and P' mark the projection onto the subspace of zero modes and the supplementary subspace, respectively.[6] Expression (24) matches the functional form of the Lipatov asymptotics (2) given above.

In accordance with relation (24), each degree of freedom for zero modes corresponds to 1/2 in the argu-

[6] In formulas (32) and (48) of [2], the power –1 of $\det[f'\{\phi_c\}]_P$ and $\det[f'\{\psi_c\}]_{P'}$ is omitted. In some cases, these determinants are functions of collective variables and should be introduced into the integral with respect to λ_i.

ment of the gamma function. This resembles the classical equipartition law. A thorough analysis reveals that this is indeed true and an exact correspondence to this law is observed. As a matter of fact, the classical partition function Z is determined by a configurational integral of $\exp(-H/T)$; as the number r_{osc} of the oscillatory degrees of freedom increases by unity, the substitution $Z \longrightarrow ZT^{1/2}$ takes place, which makes an additional contribution of 1/2 to the heat capacity [17]. Integral (14), which we are interested in, is determined by the exponential $\exp(-S\{\phi\}g)$, where the coupling constant g plays the role of temperature. An increase in the number r of zero modes by unity corresponds to a decrease in r_{osc} by unity and leads to the substitution $Z \longrightarrow Zg^{-1/2}$ (see relation (22)). In the calculation of the Lipatov asymptotics, factor $g^{-1/2}$ is estimated at the saddle point $g_{\mathrm{c}} \sim 1/N$ (see relation (19)), which leads to the substitution $Z_N \longrightarrow Z_N N^{1/2}$ and to the addition of 1/2 to the argument of the gamma function.

The equipartition law can be violated in the presence of soft modes associated with approximate symmetries of the system. In this case, some degrees of freedom in the first approximation appear as zero modes; however, in a more accurate analysis, these degrees of freedom correspond to motion in a potential relief that may be irreducible to a quadratic minimum. Examples of soft modes are dilatations in the massive four-dimensional or $(4-\epsilon)$-dimensional φ^4 theory [18, 19] and the change in the distance between elementary instantons in a two-instanton configuration (see below).

3. DISPERSION RELATION AND ROTATION RULE

In view of the factorial increase in coefficients Z_N, series (1) has zero radius of convergence. This is due to the fact that point $g = 0$ is a branching point; to single out the regular branch of $Z(g)$, we must cut the complex plane g from 0 to ∞. It is convenient to cut along the ray

$$g = |g|\operatorname{sgn} S_0, \tag{26}$$

on which the Borel sum of series (1) is poorly defined.[7] Using the Cauchy formula and writing it in the form of a dispersion relation, we obtain

$$Z(g) = \frac{1}{2\pi i}\oint_C \frac{Z(\tilde g)}{\tilde g - g}d\tilde g = \frac{1}{2\pi i}\int_0^{\infty\cdot\operatorname{sgn}S_0} \frac{\Delta Z(\tilde g)}{\tilde g - g}d\tilde g, \tag{27}$$

$$\Delta Z(g) = Z(g + i0\operatorname{sgn}S_0) - Z(g - i0\operatorname{sgn}S_0), \tag{28}$$

where contour C embraces the point $\tilde g = g$ and then is deformed so as to circumvent the cut. Expanding function (27) into a series in g, we obtain the following expression for the expansion coefficients:

$$Z_N = \frac{1}{2\pi i}\int_0^{\infty\cdot\operatorname{sgn}S_0} \frac{\Delta Z(g)}{g^{N+1}}dg. \tag{29}$$

The asymptotic form of Z_N for $N \longrightarrow \infty$ and the discontinuity on the cut for $g \longrightarrow 0$ are connected via the relation

$$\begin{aligned}\Delta Z(g) &= 2\pi i c\left(\frac{S_0}{g}\right)^b \exp\left(-\frac{S_0}{g}\right)\\ \longleftrightarrow\quad Z_N &= cS_0^{-N}\Gamma(N+b),\end{aligned} \tag{30}$$

which can easily be derived from formula (29) or by calculating the discontinuity in the cut of the Borel sum of series (1). The next step is to identify $\Delta Z(g)$ with the result of the saddle-point evaluation of integral $Z(g)$ in the vicinity of the same configuration φ_{c} as in the Lipatov method,

$$\Delta Z(g) = [Z(g)]_{\text{saddle–point }\phi_{\mathrm{c}}}. \tag{31}$$

Relations (30) and (31) were proposed by Bogomolny [13] and Parisi [14] and form the basis of the approach to calculating higher orders, which is an alternative to the Lipatov method. These relations enable us to easily find the asymptotic form of Z_N if the result of the saddle-point calculation of $Z(g)$ is already known. Relation (31) can be substantiated for a conventional integral using the elegant analysis proposed by Langer [12, 20]; however, this relation has never been proved in general form; besides, it is poorly defined and requires an appropriate interpretation.

To substantiate relation (31), we must formulate the rule of permutation of integrations with respect to g and φ; we introduce this rule using a conventional integral as an example:

$$Z(g) = \int_0^\infty d\varphi \exp(-\varphi^2 - g\varphi^4), \tag{32}$$

$$Z_N = \int_C \frac{dg}{2\pi i g}\int_0^\infty d\varphi \exp(-\varphi^2 - g\varphi^4 - N\ln g). \tag{33}$$

Diverging series (1) is obtained as a result of a regular expansion of the exponent in relation (32) in g followed by interchanging summation and integration; however, the direct expansion of $Z(g)$ into a series is not quite correct since it corresponds to the Taylor expansion at a definitely singular point. For this reason, it is appropriate in expression (33) to integrate first with respect to g (determining the coefficient of the expansion of $\exp(-g\varphi^4)$) and then with respect to φ; the con-

[7] In the φ^4 theory, action S_0 is smaller than zero and the cut is made along the negative semiaxis.

tour C can be chosen circular so that it passes through the saddle point $g_c = -1/N$ in the vertical direction (Fig. 2a). Interchanging the integration sequence is not quite trivial in view of the presence of the cut in $Z(g)$ along the ray (26). We confine the integration to the domain of large φ values using a cutting multiplier to shift the cut by a distance Δ from zero and deform the contour as shown in Fig. 2b. If we expand the contour and remove the truncation, the cut will be circumvented in the negative direction; it can easily be seen that this corresponds to formula (29).[8]

Depending on the sequence of integration, the saddle point g_c is passed either in the vertical (see Fig. 2a) or in the horizontal direction (Fig. 2b). It is well known that we should pass by this point in the steepest descend direction (it remains to be seen whether this direction is vertical or horizontal).

In integral (33), the saddle point takes place at $\varphi_c = \sqrt{2N}$, $g_c = -1/N$ and the quadratic form appearing in the expansion in the vicinity of this point ($\varphi = \varphi_c + \delta\varphi$, $g = g_c + itg_c\sqrt{2/N}$) can be written as

$$-(\delta\varphi)^2 - (t - i\sqrt{2}\delta\varphi)^2 \quad \text{or} \quad t^2 + (\delta\varphi + i\sqrt{2}t)^2. \quad (34)$$

If we integrate first with respect to g, the shift in variable t leads to the well-defined Gaussian integral

$$\int_{-\infty}^{\infty} d\varphi \int_{-\infty}^{\infty} dt \exp[-(\delta\varphi)^2 - (t - i\sqrt{2}\delta\varphi)^2] = \int_{-\infty}^{\infty} d\varphi \int_{-\infty}^{\infty} dt \exp[-(\delta\varphi)^2 - t^2]. \quad (35)$$

If we integrate first with respect to φ, we obtain a "poor" integral

$$\int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\varphi \exp[(\delta\varphi + i\sqrt{2}t)^2 + t^2] = \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\varphi \exp[(\delta\varphi)^2 + t^2]. \quad (36)$$

Let us turn the contours of integration with respect to t and φ in Eq. (36) through the same angle in the opposite directions ($t \longrightarrow te^{-i\alpha}$, $\delta\varphi \longrightarrow \delta\varphi e^{i\alpha}$); this does not change the determinant of the quadratic form, which defines the Gaussian integral. For $\alpha = \pi/2$, integral (36) is transformed into (35) and a transition from Fig. 2c to Fig. 2d takes place: the vertical integration with respect to g becomes horizontal, as is required by formula (29), while the discontinuity at the cut $\Delta Z(g)$ is obtained from the initial integral $Z(g)$ by rotating the contour relative to the saddle point φ_c through an angle $\pi/2$ in the positive direction.

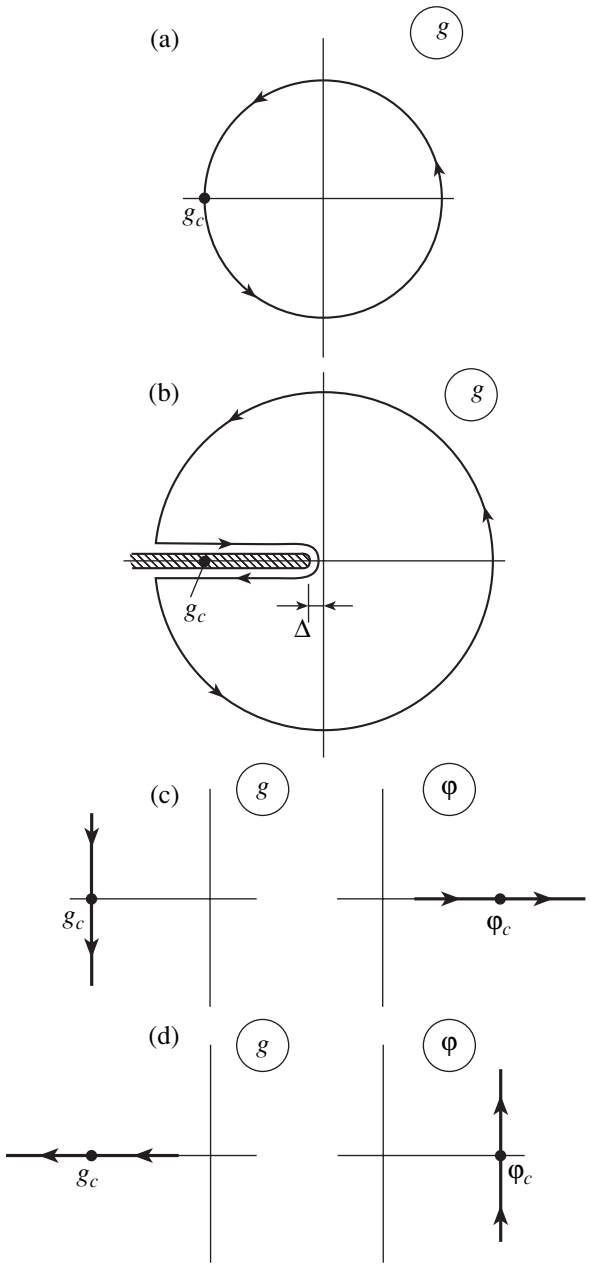


Fig. 2. (a) The contour of integration with respect to g in expression (33) can be chosen in the form of a circle if we first integrate with respect to g; (b) contour of integration with respect to g after the change in the sequence of integration; (c, d) contours of integration with respect to g and φ are rotated in the opposite directions upon the change in the order of integration.

[8] At first glance, another line of reasoning is possible. First, contour C should be deformed so that it embraces the negative semiaxis and the cut appears within it (this ensures the circumvention of the cut in the positive direction). We mean that the cut "grows" from zero to infinity. This corresponds to the Taylor expansion at the singular point and, hence, is incorrect.

The "rotation rule" formulated above can easily be extended to the general case. The main contribution to integral (18) comes from the configuration of ϕ_c corresponding to the maximum of the integrand. Consequently, the quadratic form in the exponential in Eq. (20) must lead to the sum of squares, which is possible for

$$-N\det S''\{\phi_c\} > 0. \tag{37}$$

However, integral (16) for $Z(g)$ in this case is poorly defined: it contains a "poor" Gaussian integration, which should be understood in the sense

$$\int_{-\infty}^{\infty} dx e^{x^2} \longrightarrow \int_{-\infty}^{\infty} dx e^{-x^2}, \tag{38}$$

after which integral (16) defines the discontinuity at the cut $\Delta Z(g)$. "Poor" integration can be regarded as unique due to a possible simultaneous reversal of signs of the squares of two variables. Thus, the square root of the determinant in Eq. (23) should obviously be assumed to be treated as

$$(\det[S''\{\phi_c\}]_{P'})^{-1/2} \longrightarrow i\left|\det[S''\{\phi_c\}]_{P'}\right|^{-1/2}, \tag{39}$$

after which the results (22) and (24) satisfy relation (30).

The rotation rule solves the problem of interpreting the discontinuity at the cut $\Delta Z(g)$ in the framework of the Gaussian approximation, while additional analysis is required in the presence of soft modes (Section 6).

4. RELATION BETWEEN CORRECTIONS TO THE ASYMPTOTICS AND HIGHER ORDER INSTANTONS

Separating the Lipatov asymptotics from coefficients Z_N, we obtain, in accordance with expression (3),

$$Z_N = cS_0^{-N}\Gamma\left(N + \frac{M+r}{2}\right)F\left(\frac{1}{N}\right), \tag{40}$$

and coefficients A_K can be expressed in terms of function $F(\epsilon)$:

$$A_K = \int_C \frac{d\epsilon}{2\pi i}\frac{F(\epsilon)}{\epsilon^{K+1}}. \tag{41}$$

Substituting expression (18) into (40), setting $\epsilon = 1/N$, and carrying out the substitution $g \longrightarrow \epsilon S_0 g$, we obtain an exact expression for A_K:

$$A_K = \frac{1}{\sqrt{2\pi}c}\int\frac{d\epsilon}{2\pi i\epsilon}\int\frac{dg}{2\pi i g}\int D\varphi\varphi^{(1)}\ldots\varphi^{(M)}\epsilon^{(M+r-1)/2} \times \exp\left\{\frac{1}{\epsilon}\left[1 - \ln g - \frac{S\{\phi\}}{S_0 g}\right] - K\ln\epsilon\right\}_{\phi = \varphi\sqrt{\epsilon g S_0}}, \tag{42}$$

which can be estimated by the saddle-point method for large values of K. The saddle-point configuration is defined by the conditions

$$S'\{\psi_c\} = 0, \quad g_c = \frac{S'\{\psi_c\}}{S_0}, \quad \epsilon_c = \frac{\ln g_c}{K}, \tag{43}$$

after the expansion in the vicinity of this configuration, the exponential assumes the form

$$\exp\left\{-\frac{1}{2}\left[\delta\varphi, S''\{\psi_c\}\delta\varphi - \frac{t^2}{\epsilon_c} - K\tau^2\right]\right\}, \tag{44}$$

where $g - gc = ig_c t$ and $\epsilon - \epsilon_c = i\epsilon_c\tau$. For ψ_c, we cannot use the configurations $\psi_c = 0$ and $\psi_c = \phi_c$ since $g_c = 0$ or $\epsilon_c = 0$ in this case; this corresponds to a singularity rather than a saddle point. The saddle point corresponds to the maximum of the integrand under the condition

$$\frac{K}{\epsilon_c}\det S''\{\psi_c\} > 0, \tag{45}$$

and we must take the first of higher order instantons for which $\det S''\{\psi_c\} > 0$ for ψ_c. This condition is usually satisfied even for the second instanton, which will be presumed in the subsequent analysis. Evaluation of integral (42) gives formula (4), in which

$$\tilde{c} = \frac{(\ln S_1/S_0)^{(r-r')/2}S_1^{-(M+r')/2}}{c(2\pi)^{2+r'/2}\det[f'\{\psi_c\}]_P} \times \sqrt{\frac{\det S''\{0\}}{\det[S''\{\psi_c\}]_{P'}}}\int\prod_{i=1}^{r'} d\lambda_i\psi_\lambda^{(1)}\ldots\psi_\lambda^{(M)}, \tag{46}$$

and $S_1 = S\{\psi_c\}$ (see [2] for details).

The dispersion relation (see Section 3) enables us to establish the relation between the corrections to the asymptotics and higher order instantons in the general form without resorting to specific features of the φ^4 theory. Let coefficients A_K increase for large values of K according to the factorial law

$$A_K = \tilde{c}\tilde{a}^K\Gamma(K + \tilde{b}) \tag{47}$$

with $\tilde{a} > 0$; in this case, the Borel sum of the series in Eq. (3) is poorly defined for $N > 0$ and the coefficient

function Z_N has a cut with a discontinuity in it defined by rule (30):

$$\Delta Z_N = cS_0^{-N}\Gamma(N+b)\times 2\pi i\tilde{c}\left(\frac{N}{\tilde{a}}\right)^{\tilde{b}}\exp\left(-\frac{N}{\tilde{a}}\right). \quad (48)$$

On the other hand, this discontinuity can be defined in analogy with (31) as the contribution of the saddle-point configuration ψ_c to the Lipatov integral (18). Assuming that the functional form of this contribution is analogous to (2),

$$\Delta Z_N = ic_1 S_1^{-N}\Gamma(N+b_1), \quad (49)$$

and identifying expression (48) with (49), we obtain

$$\tilde{a} = \left(\ln\frac{S_1}{S_0}\right)^{-1}, \quad \tilde{b} = b_1 - b, \quad \tilde{c} = \frac{c_1}{2\pi c}\left(\ln\frac{S_1}{S_0}\right)^{b-b_1}. \quad (50)$$

As a result, the determination of the parameters of the asymptotic form of A_K can be reduced to the well-known procedure: it is sufficient to evaluate the saddle-point contributions (2) and (49) to the Lipatov integral (18) from two configurations, ϕ_c and ψ_c. Result (46) readily follows from (50) if we take into account the fact that expression (49) can be derived from (25) via the substitution $\phi_c \longrightarrow \psi_c$, $r \longrightarrow r'$, and $(-\det[S''\{\psi_c\}]_{P'})^{-1/2}$ is treated as $i|\det[S''\{\psi_c\}]_{P'}|^{-1/2}$ in accordance with the rotation rule.

Formulas (50) solve the problem of evaluating the asymptotic form of A_K under the condition of applicability of the equipartition law; for this purpose, all fluctuational modes in the vicinity of classical configuration ϕ_c and ψ_c must be distinctly separable into zero and vibrational modes. In this case, $b_1 - b = (r' - r)/2$ irrespective of the specific features of the φ^4 theory: a contribution of the $M/2$ type in the argument of the gamma function, which stems from the preexponential factor in formula (14), may have different values in other field theories, but it is the same for the first and second instantons. In the presence of soft modes, the situation is more complicated and will be considered in the following sections.[9]

5. INSTANTON COMBINATION RULE

Let us find out how to construct the contribution from a two-instanton configuration knowing the contribution from one instanton to the functional integral (14). In accordance with Section 2, the contribution to $Z_M(g)$ from the saddle-point configuration ϕ_c has the structure

$$Z_M^{(1)}(g) = c_0 g^{-(M+r)/2}\exp\left(-\frac{S_0}{g}\right)\int\prod_{i=1}^{r} d\lambda_i \phi_\lambda^{(1)}\ldots\phi_\lambda^{(M)}, \quad (51)$$

where $c_0 = c_0\{\phi_c\}$ and $S_0 = S\{\phi_c\}$ are functionals of ϕ_c. The contribution from the two-instanton configuration is defined by an analogous expression, in which ϕ_c is replaced by $\phi_\lambda + \phi_{\lambda'}$. If we introduce the instanton interaction $S_{\rm int}$ by the relation

$$S\{\phi_\lambda + \phi_{\lambda'}\} = S\{\phi_\lambda\} + S\{\phi_{\lambda'}\} + S_{\rm int}\{\phi_\lambda, \phi_{\lambda'}\} \quad (52)$$

and take into account the doubling of the number of collective variables, we obtain the sum of terms of the form

$$Z_{LL'} = c_1 g^{-M/2-r}\exp\left(-\frac{2S_0}{g}\right)\int\prod_{i=1}^{r} d\lambda_i d\lambda_i' \phi_\lambda^{(1)}\ldots\phi_\lambda^{(L)} \times \phi_{\lambda'}^{(1)}\ldots\phi_\lambda^{(L')}\exp\left(-\frac{S_{\rm int}\{\phi_\lambda, \phi_{\lambda'}\}}{g}\right), \quad (53)$$

where $L + L' = M$. For small values of g, the exponential limits the interaction between instantons by the condition $S_{\rm int}\{\phi_\lambda, \phi_{\lambda'}\} \lesssim g$; consequently, we can disregard this term in the preexponential factor. It can naturally be expected that c_1 must be equal to c_0^2 for expression (53) at $S_{\rm int} \equiv 0$ to be just the product of the right-hand sides of Eqs. (51) with $M = L$ and $M = L'$. Small values of $S_{\rm int}$ correspond to remote instantons,[10] which enables us to disregard cross terms simultaneously containing ϕ_λ and $\phi_{\lambda'}$. In this case, the sum over L, L' will contain only two terms with $L = M$, $L' = 0$ and $L = 0$, $L' = M$, which are obviously identical. The emerging factor 2 is cancelled out with the combinatorial multiplier 1/2!, which should be introduced in view of the fact that configurations differing in the transposition of instantons are taken into account twice. As a result, the two-instanton contribution assumes the form

$$Z_M^{(2)}(g) = c_0^2 g^{-M/2-r}\exp\left(-\frac{2S_0}{g}\right) \times \int\prod_{i=1}^{r} d\lambda_i d\lambda_i' \phi_\lambda^{(1)}\ldots\phi_\lambda^{(M)}\exp\left(-\frac{S_{\rm int}\{\lambda, \lambda'\}}{g}\right), \quad (54)$$

where the fact that $S_{\rm int}\{\phi_\lambda, \phi_{\lambda'}\}$ for a fixed formed of instantons depends only on λ and λ' is taken into account. For $M = 0$, configurations with $L = M$, $L' = 0$ and $L = 0$, $L' = M$ coincide and the result is defined by formula (54) with an additional factor 1/2 on the right-

[9] Relations of type (50) are valid in this case as well (with allowance for possible change in the functional form of expressions (47) and (49)); however, these relations are practically useless since the emerging integrals are poorly defined.

[10] See Section 7 for $d = 4$.

hand side; this remark applies to all subsequent expressions as well.

Formula (54) enables us to write the expression for the two-instanton contribution from the well-known result (51) for the one-instanton contribution; we only require additional information about the interaction of instantons at large distances. The instanton combination rule appears quite natural, but some fine aspects missing in a heuristic derivation must be discussed.

Introduction of a constraint. In fact the replacement of ϕ_c by $\phi_\lambda + \phi_{\lambda'}$ is not quite correct since the linear combination of instantons is not the exact solution to the equation $S'\{\phi\} = 0$. For this reason, the expansion in the vicinity of this configuration acquires terms which are linear in $\delta\phi$ and require accurate elimination.[11]

Let us introduce the collective variable z characterizing the separation R between the instantons and formally defined for an arbitrary instanton configuration, $z = f\{\phi\}$. The idea consists in finding the extremum of action under an additional condition (constraint) $f\{\phi\} = \text{const}$ (i.e., for a fixed distance between instantons) and in subsequent integration with respect to this distance. In this case, the instanton is defined by the equation

$$S'\{\phi_c\} - \mu f'\{\phi_c\} = 0 \tag{55}$$

(μ is the Lagrange multiplier) and the integration with respect to z is carried out by introducing the expansion of unity,

$$\begin{aligned} 1 &= \int dz\,\delta(z - f\{\phi\}) \\ &= \int dz\,\delta(z - f\{\phi_c\} - (f'\{\phi_c\}, \delta\phi)), \end{aligned} \tag{56}$$

into the functional integral. Choosing the instanton from the condition $z = f\{\phi_c\}$, we obtain

$$\begin{aligned} Z(g) &= \int D\varphi\varphi^{(1)}\ldots\varphi^{(M)} \\ &\times \exp\left\{-\frac{S\{\phi_c\} + (S'\{\phi_c\}, \delta\phi) + (\delta\phi, S''\{\phi_c\}\delta\phi)/2}{g}\right\} \\ &\times \int dz\,\delta(-(f'\{\phi_c\}, \delta\phi)) \end{aligned} \tag{57}$$

and the terms linear in $\delta\phi$ in the exponential are eliminated by the δ-function in view of condition (55). For $f\{\phi\}$, it is convenient to take the quantity $S_{\text{int}}\{\phi_\lambda, \phi_{\lambda'}\}$ since Eq. (55) has a combination of instantons $\phi_\lambda + \phi_{\lambda'}$ as the exact solution for $\mu = 1$ (cf. formula (52)).

We can disregard the instanton interaction in the preexponential factor; in this case, zero modes are separated as if they were independent. The only fine point is that, instead of zero modes

$$\frac{\partial\phi_\lambda}{\partial x_1} \text{ and } \frac{\partial\phi_{\lambda'}}{\partial x_1}, \tag{58}$$

corresponding to translations of instantons along the straight line passing through their centers and coinciding with the x_1 axis, we must take their linear combinations

$$\frac{\partial\phi_\lambda}{\partial x_1} + \frac{\partial\phi_{\lambda'}}{\partial x_1} \text{ and } \frac{\partial\phi_\lambda}{\partial x_1} - \frac{\partial\phi_{\lambda'}}{\partial x_1}, \tag{59}$$

corresponding to translation of a pair as a whole and to a change in the spacing between the pair components. In this case, the δ-function corresponding to the second mode in (59) is not introduced into product (21) since this role is played by expansion (56). This modification is of no importance since $\det[f'\{f_c\}]_P$ in expressions (23) and (25) is in fact determined by the Gram matrix constructed on zero modes (Section 8) and is independent of the choice of functionals $f_i\{\varphi\}$. Consequently, the final result (44) corresponds to the formal substitution $\phi_c \longrightarrow \phi_\lambda + \phi_{\lambda'}$.

Factorization of determinants. While deriving formula (54), we assumed that $c_1 = c_0^2$, or

$$c_0\{\phi_\lambda + \phi_{\lambda'}\} = c_0\{\phi_\lambda\}c_0\{\phi_{\lambda'}\}. \tag{60}$$

To clarify this relation, we note that operator $S''\{\phi_c\}$ in the scalar case has the form of the Schrödinger operator (Section 8)

$$S''\{\phi_c\} = -\Delta + m^2 - 3\phi_c^2(x), \tag{61}$$

and, instead of a single potential well, two potential wells separated by a large distance appear upon the substitution $\phi_c \longrightarrow \phi_\lambda + \phi_{\lambda'}$. Consequently, the eigenvalues of operator $S''\{\phi_\lambda + \phi_{\lambda'}\}$ are doubly degenerate eigenvalues of operator $S''\{\phi_c\}$ and

$$\det S''\{\phi_\lambda + \phi_{\lambda'}\} = [\det S''\{\phi_c\}]^2 \tag{62}$$

under the condition that the product of eigenvalues converges. No such convergence is observed for operator $S''\{\phi_c\}$, and the normalization of integral (14) to quantity $Z_0(0)$ is significant. As a result of this normalization, the combination

$$\frac{\det S''\{\phi_c\}}{\det S''\{0\}} = \det\frac{S''\{\phi_c\}}{S''\{0\}} \tag{63}$$

appears, where operator $S''\{\phi_c\}/S''\{0\}$ has a discrete spectrum [21] and the convergence of the products of eigenvalues is ensured by a simple renormalization (see [4] for details). The multipliers singled out from

[11] The instanton interaction depends to a considerable extent on the specific choice of the two-instanton configuration; with an inappropriate choice, the results can easily be erroneous (cf. [8–12]).

relation (63) as a result of renormalization and during the elimination of zero modes as a result of the substitution $\phi_c \longrightarrow \phi_\lambda + \phi_{\lambda'}$ are squared in view of the equivalence of two instantons. Factorization of $\det[f'\{\phi_\lambda + \phi_{\lambda'}\}]_P$ similar to (62) is due to the fact that zero modes are considered in the approximation of noninteracting instantons.

It should be noted that factorization (60) is not valid in the combination of topological instantons connecting degenerate but nonequivalent vacuums [8, 12]. In this case, the effective potential appearing in the Schrödinger operator of type (61) does not appear as two isolated potential wells: a potential barrier emerging between the wells renders the interaction between the instanton a long-range interaction.

Subtraction of the contribution from an ideal gas. As we pass to the Green function, result (54) assumes the form

$$G_M^{(2)}(g) = c_0^2 g^{-M/2-r}\exp\left(-\frac{2S_0}{g}\right) \times \int\prod_{i=1}^{r} d\lambda_i d\lambda_i' \phi_\lambda^{(1)}\ldots\phi_\lambda^{(M)}\left[\exp\left(-\frac{S_{\text{int}}\{\lambda,\lambda'\}}{g}\right)-1\right]; \quad (64)$$

i.e., the contribution of noninteracting instantons is subtracted in the same way as in the case of the virial expansion [17]. Indeed, we can write the Green function in the form

$$G_M(g) = \frac{Z_M(g)}{Z_0(g)} = \frac{Z_M^{(0)}(g) + Z_M^{(1)}(g) + Z_M^{(2)}(g) + \ldots}{1 + Z_0^{(1)}(g) + Z_0^{(2)}(g) + \ldots}, \quad (65)$$

where the superscripts "0," "1," "2," … mark the contributions from saddle-point configurations[12] $\phi = 0$, $\phi = \phi_c$, $\phi = \phi_\lambda + \phi_{\lambda'}$, … and having the order 1, $\exp(-S_0/g)$, $\exp(-2S_0/g)$, … . For convenience, the normalization of functional integrals in this expression is carried out not to $Z_0(0)$, but to $Z_0^{(0)}(g)$, which is immaterial in the main order in g. The contribution to $G_M(g)$ on the order of $\exp(-2S_0/g)$ has the form

$$Z_M^{(2)}(g) - Z_M^{(1)}(g)Z_0^{(1)}(g) - Z_M^{(0)}(g)Z_0^{(2)}(g), \quad (66)$$

where the last term is small for $M > 0$ in keeping with the smallness of $g^{M/2}$. It can easily be seen that the second term corresponds to expression (54) with $S_{\text{int}} \equiv 0$, which gives –1 in formula (64). Similarly, after taking the logarithm of the vacuum integral,

$$\ln Z_0^{(0)}(g) = \ln[1 + Z_0^{(1)}(g) + Z_0^{(2)}(g) + \ldots] = Z_0^{(1)}(g) + Z_0^{(2)}(g) - \frac{1}{2}[Z_0^{(1)}(g)]^2 + \ldots, \quad (67)$$

the last term leads to the subtraction of the contribution from noninteracting instantons in the expression for $Z_0^{(2)}(g)$; in this case, the presence of the additional factor 1/2 is significant in expressions with $M = 0$.

6. HIGHER ORDER CORRECTIONS TO ASYMPTOTICS IN THE PRESENCE OF SOFT MODES

Let us analyze expression (64) in the case of a power interaction between instantons separated by large distances:

$$S_{\text{int}}\{\lambda,\lambda'\} = \frac{a(\lambda,\lambda')}{R^\alpha}. \quad (68)$$

Introducing into expression (64) the expansion of unity,

$$1 = \int_0^\infty dz\,\delta\left(z - \frac{S_{\text{int}}\{\lambda,\lambda'\}}{2S_0}\right) + \int_0^\infty dz\,\delta\left(z + \frac{S_{\text{int}}\{\lambda,\lambda'\}}{2S_0}\right), \quad (69)$$

and considering that integration with respect to collective variables includes the integral $\int R^{d-1}dR$, we obtain

$$\int\prod_{i=1}^{r}\delta\lambda_i\delta\lambda_i'\phi_\lambda^{(1)}\ldots\phi_\lambda^{(M)}\delta\left(z \pm \frac{S_{\text{int}}\{\lambda,\lambda'\}}{2S_0}\right) = \frac{A^\pm}{z^{1+\nu}}, \quad (70)$$

where $\nu = d/\alpha$. As a result, expression (64) assumes the functional form

$$G^{(2)}(g) = B\left(\frac{2S_0}{g}\right)^{(M+2r)/2}\exp\left(-\frac{2S_0}{g}\right)F\left(\frac{2S_0}{g}\right), \quad (71)$$

where

$$F(x) = A^+I^+(x) + A^-I^-(x), \quad (72)$$

$$I^\pm(x) = \int_0^\infty \frac{dz}{z^{1+\nu}}(e^{\pm xz} - 1). \quad (73)$$

[12] Such an "expansion in instantons" in the case of an ordinary integral is obtained by deforming the integration contour in such a way that it can be presented as the sum of integrals over contours C_0, C_1, C_2, …, where each contour C_i passes through the saddle point z_i in the steepest descent direction and its ends pass to infinity.

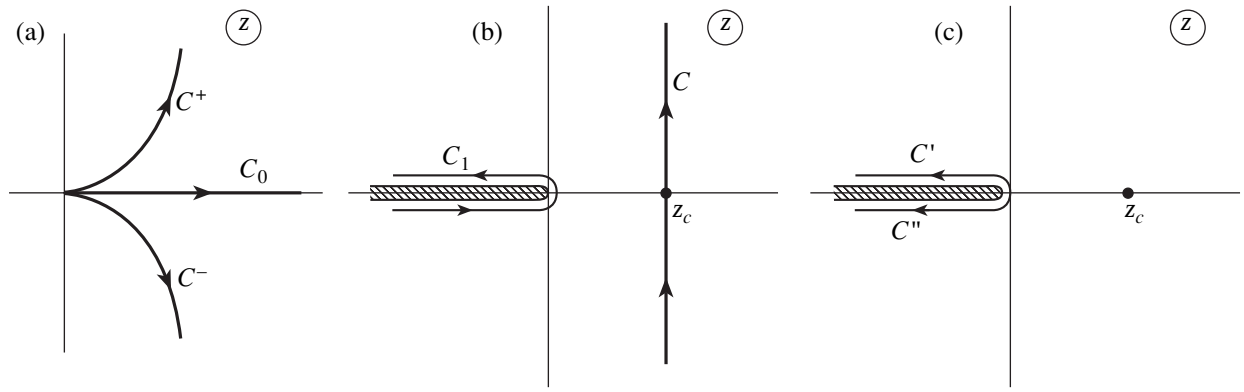


Fig. 3. (a) Integration contour C_0 in integral $I^+(x)$ is bent upwards or downwards upon a displacement of g to the complex plane; (b) integration contour C defines the discontinuity at the cut of integral $I^+(x)$; to evaluate the integral, the contour is deformed to C_1; (c) in the calculation of the asymptotic form of A_K, half-sum of the integrals over contours C' and C'' appears.

We will henceforth assume that $\nu \geq 0$ since negative values of ν correspond to nonphysical increase in the interaction intensity with the distance. For $\nu \geq 1$, subtraction of unity does not ensure the convergence of the integral in formula (73). Such values of ν correspond to a slowly decreasing interaction, for which the virial expansion is not applicable. The interaction for large values of R should be modified in the spirit of Debye screening, which leads to truncation of the integral for small values of z. The type of truncation is immaterial since integral $I^-(x)$ can be evaluated by differentiating with respect to parameter x; for $F(x)$, we formally get

$$F(x) = \Gamma(-\nu)[A^- x^\nu + A^+(-x^\nu) + O(x^{[\nu]})], \qquad (74)$$

where $[\ldots]$ is the integral part of the number. The values of g near the cut, where $x = 2S_0/g > 0$ and expression (74) is poorly defined in view of integral $I^+(x)$ being poorly defined is of practical importance. The interpretation of this expression depends on the formulation of the problem; two versions of the interpretation will be considered below.

Contribution to the Lipatov asymptotics. Analysis of the contribution from a two-instanton configuration to the asymptotic form of coefficients Z_N is important for problems with degenerate vacuum, in which a solitary instanton is topological and the Lipatov asymptotics is determined by the instanton–antiinstanton pair [8–12]. In view of formula (71), the discontinuity at the cut of function $G(g)$ is defined by the formula

$$\Delta G(g) = \pm B\left(\frac{2S_0}{g}\right)^{(M+2r)/2} \exp\left(-\frac{2S_0}{g}\right) \times \left[F\left(\frac{2S_0}{g} - i0\right) - F\left(\frac{2S_0}{g} + i0\right)\right], \qquad (75)$$

where we represent $G(g)$ as an "expansion" in instantons $G^{(0)}(g) + G^{(1)}(g) + G^{(2)}(g) + \ldots$ (see Section 4) and assume that $G^{(0)}(g)$ and $G^{(1)}(g)$ make zero contributions to the discontinuity at the cut. The indefinite sign in formula (75) is due to the fact that the expansion in instantons requires a preliminary deformation of the integration contour so that it passes through all saddle points (see Footnote 12): in the general case, such a deformation is ambiguous and leads to the indefinite sign of $G^{(2)}(g)$ in view of the possibility to pass through a saddle point in two opposite directions. It can easily be seen that the well-defined integral $I^-(x)$ makes zero contribution to discontinuity (75), while in expression for $I^+(x)$, we can omit the term with –1. As g is shifted to the lower or upper half-plane, the contour of integration with respect to z in integral $I^+(x)$ should be bent upwards or downwards (Fig. 3a). The discontinuity at the cut is determined by the difference in the integrals over contours C^+ and C^- and can be reduced to the vertical contour C (Fig. 3b).

To fix the sign in expression (75), we must establish its relation with the rotation rule (see Section 3). We replace $R^{-\alpha}$ in Eq. (68) by $R^{-\alpha} + \epsilon R^\alpha$; then the exponent in integral $I^+(x)$ acquires $x(z + \epsilon/z)$ and a saddle point $z_c = \sqrt{\epsilon}$ appears, through which contour C should be drawn. The integration must be carried out in the upward direction so that contour C is obtained from the initial contour C_0 by rotation through $\pi/2$ in the positive direction around point z_c. For $\epsilon \longrightarrow 0$, the saddle point disappears, but the direction of contour C is preserved and corresponds to the negative sign in expression (75). Evaluating integral $I^+(x)$ using the deformation of contour C to position C_1, we obtain

$$I^+(x) \longrightarrow \frac{2\pi i}{\Gamma(1+\nu)} x^\nu, \qquad (76)$$

which gives the following expression for the disconti-

nuity at the cut of function $G(g)$:

$$\Delta G(g) = 2\pi i \frac{BA^+}{\Gamma(1+\nu)}\left(\frac{2S_0}{g}\right)^{M/2+r+\nu}\exp\left(-\frac{2S_0}{g}\right). \quad (77)$$

The Lipatov asymptotics for G_N is obtained in accordance with the correspondence rule (30),

$$G_N = \frac{BA^+}{\Gamma(1+\nu)}(2S_0)^{-N}\Gamma(N+M/2+r+\nu), \quad (78)$$

and is well defined for all values of $\nu > -1$. Expression (76) corresponds to the algorithm correctly formulated by Bogomolny and Fateyev [9], although their line of reasoning leads to a result with the opposite sign.[13]

Asymptotic form of coefficients A_K. Analogously to expression (42), we have the following exact expression for coefficients A_K corresponding to quantity $G(g)$:

$$A_K = \frac{1}{\sqrt{2\pi}c}\int\frac{d\epsilon}{2\pi i\epsilon}\int\frac{dg}{2\pi i g}\epsilon^{(M+r-1)/2}G(g) \times \exp\left\{\frac{1}{\epsilon}\left(\frac{S_0\epsilon}{g}+1\right)-K\ln\epsilon\right\}. \quad (79)$$

Substituting quantity $G^{(2)}(g)$ from Eq. (71) for $G(g)$ and evaluating the integrals with respect to g and ϵ in the saddle-point approximation, we obtain

$$A_K = \pm\frac{B}{c(2\pi)^2(\ln 2)^{r/2}} \times \Gamma\left(K+\frac{r}{2}\right)(\ln 2)^{-K}F\left(\frac{K}{\ln 2}\right). \quad (80)$$

Vertical contours of integration with respect to g and ϵ correspond to "poor" Gaussian integrals (cf. formula (44)) and must be rotated simultaneously to the horizontal position in accordance with Section 3. The sign in formula (80) turns out to be indeterminate for the same reason as in the above case. To fix the sign, we replace $R^{-\alpha}$ by $R^{-\alpha} + \epsilon R^{\alpha}$ and consider the saddle-point configuration in the range of parameters, where $S_{\rm int} > 0$, which corresponds to integral $I^-(x)$. For this region, $\det S''\{\psi_c\} > 0$ and the sign in formula (80) must correspond to the arithmetic root of the determinant, which leads to the condition $\pm B > 0$. Formula (80) is valid if function $F(x)$ is well defined on the real axis. Otherwise, $F(x)$ should be treated as the half-sum of $F(x + i0)$ and $F(x - i0)$ since half the saddle-point contribution stems from region $\mathrm{Im}\, g > 0$, while the other half, from region $\mathrm{Im}\, g < 0$. Thus, we can write

$$A_K = \frac{|B|}{c(2\pi)^2(\ln 2)^{r/2}}\Gamma\left(K+\frac{r}{2}\right) \times (\ln 2)^{-K}\mathrm{Re}F\left(\frac{K}{\ln 2}+i0\right) \quad (81)$$

and the substitution of Eq. (74) gives

$$A_K = \frac{|B|\Gamma(-\nu)(A^- + A^+\cos\pi\nu)}{c(2\pi)^r(\ln 2)^{\nu+r/2}} \times \Gamma\left(K+\frac{r}{2}+\nu\right)(\ln 2)^{-K}. \quad (82)$$

The poorly defined integral $I^+(x)$ is treated as the half-sum of integrals over contours C' and C'' (see Fig. 3c),

$$I^+(x) \longrightarrow \frac{1}{2}[I^+(x+i0) + I^+(x-i0)] = \Gamma(-\nu)x^{\nu}\cos\pi\nu \quad (83)$$

(cf. formula (76)). If the saddle point is created artificially by replacing $R^{-\alpha}$ by $R^{-\alpha} + \epsilon R^{\alpha}$, contours C' and C'' correspond to the steepest decent as before and the saddle point z_c makes zero contribution to A_K in accordance with the fact that $\det S''\{\psi_c\} < 0$ for this point (Section 4). It can be seen that for a power law of instanton interaction (68), the argument of the gamma-function acquires an additional contribution $\nu = d/\alpha$ as compared to formula (4). This contribution is in fact important only for $d = 4$. For $d < 4$, the interaction of instantons is exponential, which corresponds to $\alpha = \infty$ and $\nu = 0$: the argument of the gamma-function matches Eq. (4), but additional logarithmic multipliers appear (Section 7).

Interpretation of integral $I^+(x)$ completes the formulation of the general computational algorithm. Proceeding from the well-known expression (51) for the one-instanton contribution, we construct the contribution from two-instanton configuration (64), in which we must calculate the interaction of instantons $S_{\rm int}(\lambda, \lambda')$ in the region where this interaction is weak. Introducing the slowly varying function $F(x)$ in accordance with formula (71), we can straightaway write the result (81).

7. COEFFICIENTS A_K IN THE φ^4 THEORY

7.1. Interaction of Instantons

Action (5) is reduced to the form (15) using the substitution[14] $\varphi_\alpha(x) = (-g)^{-1/2}\phi_\alpha(x)$. An instanton is pre-

[13] The incorrect sign also appeared in the work by Balitsky [22], who calculated the Lipatov asymptotics in QCD and was corrected in the review by Zakharov [23] proceeding from physical considerations.

[14] The additional minus as compared to relation (15) ensures the real-valuedness of ϕ_c. In this case, the substitutions $g^{-b} \longrightarrow (-g)^{-b}$ and $S_0^{-N} \longrightarrow (-S_0)^{-N}$ take place in Eqs. (22) and (24).

sented in the form $\phi_\alpha(x) = u_\alpha\phi_c(x)$, where u_α are the components of the unit vector $\mathbf{u}$ and $\phi_c(x)$ is the solution to the equation

$$-\Delta\phi + m^2\phi - \phi^3 = 0. \quad (84)$$

Using the general form of the two-instanton configuration,

$$\psi_\alpha(x) = u_\alpha\phi_c(x) + u'_\alpha\phi_c(x+R) \equiv u_\alpha\phi + u'_\alpha\phi_R, \quad (85)$$

for the interaction of instantons, we have

$$S_{\rm int}\{\mathbf{u}, \mathbf{u}', R\} = \int d^dx \times\left\{(\mathbf{u}\mathbf{u}')\phi^3\phi_R + (\mathbf{u}\mathbf{u}')^2\phi^2\phi_R^2 + \frac{1}{2}\phi^2\phi_R^2\right\}, \quad (86)$$

where we have used definition (52) and eliminated gradients with the help of Eq. (84). For $d < 4$, we can assume that $m = 1$, since the substitution $x \longrightarrow x/m$, $\phi \longrightarrow \phi m$ eliminates the mass from Eq. (84), after which it disappears in all dimensionless quantities. For large values of $|x|$, the spherically symmetric solution to Eq. (84) has the form

$$\phi_c(x) = \text{const}\cdot|x|^{-\mu}K_\mu(|x|) = \text{const}\cdot|x|^{(1-d)/2}e^{-|x|}, \quad (87)$$

where $\mu = (d-2)/2$ and $K_\mu(x)$ is the Macdonald function. The substitution of relation (87) into Eq. (86) shows that the main contribution is determined by the first term,

$$S_{\rm int}\{\mathbf{u}, \mathbf{u}', R\} \approx (\mathbf{u}\cdot\mathbf{u}')\phi_c(R+\bar{x})\int\phi_c^3(x)d^dx \approx \text{const}\cdot(\mathbf{u}\cdot\mathbf{u}')R^{(1-d)/2}e^{-R}, \quad (88)$$

where $\bar{x} \sim 1$. For $d = 1$, the constant can be evaluated using the explicit form of instanton $\phi_c(x) = \sqrt{2}/\cosh x$,

$$S_{\rm int}\{\mathbf{u}, \mathbf{u}', R\} = 16(\mathbf{u}\cdot\mathbf{u}')e^{-R}. \quad (89)$$

In the four-dimensional case, the massless theory is of principal importance, in which Eq. (84) has a solution

$$\phi_c(x) = \frac{2\sqrt{2}\rho}{x^2+\rho^2} \quad (90)$$

with an arbitrary instanton radius ρ. In the linear combination (85), we must permit various radii ρ and ρ_1 for functions ϕ and ϕ_R. Disregarding quantity x in relation (88), we obtain

$$S_{\rm int}\{\mathbf{u}, \mathbf{u}', \rho, \rho_1, R\} = 32\pi^2(\mathbf{u}\cdot\mathbf{u}')\frac{\rho\rho_1}{R^2+\rho_1^2}, \quad (91)$$

where we assume that

$$\rho \ll \rho_1 \sim R, \quad (92)$$

since precisely these configurations are of interest for the subsequent analysis.[15]

7.2. Results for d < 4

Proceeding from the well-known Lipatov asymptotics (formula (79) in [18]) and using the correspondence rule (30), for the one-instanton contribution, we obtain

$$Z_M^{(1)}(\alpha_1x_1, \ldots, \alpha_Mx_M) = ic_0(-g)^{-(M+r)/2}\exp(-S_0/g) \times\int d^dx_0\phi_c(x_1-x_0)\ldots\phi_c(x_M-x_0) \times\int d^nu\,\delta(|u|-1)u_{\alpha_1}\ldots u_{\alpha_M}, \quad (93)$$

where

$$S_0 = -\frac{I_4}{4}, \quad r = n-1+d, \quad I_p = \int d^dx\phi_c^p(x),$$

$$c_0 = \frac{1}{(2\pi)^{r/2}}\left(\frac{I_6-I_4}{d}\right)^{d/2} \times I_4^{(n-1)/2}\left[-\bar{D}_R(1)\bar{D}_R^{n-1}\left(\frac{1}{3}\right)\right]^{-1/2}; \quad (94)$$

$\phi_c(x)$ is the solution to Eq. (84) with $m = 1$; and $\bar{D}_R(1)$ and $\bar{D}_R(1/3)$ are the renormalized determinants whose values will be given below. In accordance with the instanton combination rule, we write the expression for the two-instanton contribution

$$G_M^{(2)}(\alpha_1x_1, \ldots, \alpha_Mx_M) = -c_0^2(-g)^{-(M+2r)/2} \times\exp\left(-\frac{2S_0}{g}\right)H(g)\int d^dx_0\phi_c(x_1-x_0)\ldots\phi_c(x_M-x_0) \times\int d^nu\,\delta(|\mathbf{u}|-1)u_{\alpha_1}\ldots u_{\alpha_M},$$

[15] The possibility of disregarding the crossed terms in Eq. (53) in this case is associated not with the large distance between the instantons, but with different degrees of their localization ϕ_λ and $\phi_{\lambda'}$.

$$H(g) = \int d^n u' \delta(|\mathbf{u}'| - 1) \times \int d^d x_0' \left\{ \exp\left[-\frac{S_{\text{int}}(\mathbf{u}, \mathbf{u}', x_0' - x_0)}{g} \right] - 1 \right\}. \quad (95)$$

The last integral ($R \equiv x_0' - x_0$),

$$\int d^d R \left(\exp\left\{ -\frac{A R^{(1-d)/2} e^{-R}}{g} \right\} - 1 \right) \approx -\frac{\sigma_d}{d} \left(\ln \left| \frac{1}{g} \right| \right)^d, \quad (96)$$

is independent of A to logarithmic accuracy (σ_d is the area of a unit sphere in the d-dimensional space); for this reason, the dependence on $\mathbf{u}'$ is absent and $H(g)$ can be obtained from relation (96) simply by multiplying it by σ_n. Following the algorithm described in Section 6, we obtain the following result for A_K:

$$A_K = -c_0 \frac{2^{-M/2}}{2\pi} (I_4 \ln 2)^{-r/2} \frac{\sigma_n \sigma_d}{d} \times \Gamma\left(K + \frac{r}{2} \right) (\ln 2)^{-K} \ln^d K. \quad (97)$$

Using the numerical values [21]

$$I_6 = 71.080, \quad I_4 = 23.402, \quad -\bar{D}_R(1) = 135.3, \quad \bar{D}(1/3) = 1.465 \quad (98a)$$

for $d = 2$ and

$$I_6 = 659.87, \quad I_4 = 75.589, \quad -\bar{D}_R(1) = 10.544, \quad \bar{D}(1/3) = 1.4571 \quad (98b)$$

for $d = 3$, we obtain formulas (9) and (10) given in the Introduction. For $d = 1$, the result can be obtained not with a logarithmic, but with a power accuracy in $1/K$. Integral (96) for $\mathbf{u} \cdot \mathbf{u}'/g > 0$ has the form

$$\sigma_1 \int_0^\infty dR \left(\exp\left\{ -\frac{16(\mathbf{u} \cdot \mathbf{u}') e^{-R}}{g} \right\} - 1 \right) = -2 \left(\ln \frac{16(\mathbf{u} \cdot \mathbf{u}')}{g} + C_{\text{E}} \right), \quad (99)$$

and integration with respect to $\mathbf{u}'$ gives

$$\text{Re} H(g + i0) = -2\sigma_n \left(\ln \frac{16}{g} + C_{\text{E}} + \frac{\psi(1/2) - \psi(n/2)}{2} \right). \quad (100)$$

Passing to the expression for A_K and using the values of parameters [4]

$$I_4 = \frac{16}{3}, \quad I_6 = \frac{128}{15}, \quad \bar{D}_R(1) = -\frac{1}{5}, \quad \bar{D}_R\left(\frac{1}{3} \right) = \frac{1}{3}, \quad (101)$$

we obtain result (8).

7.3. Four-Dimensional Case

In the four-dimensional case, the expression for the one-instanton contribution differs from (93) in view of the presence of an extra zero mode associated with possible variation of instanton radius ρ. This expression can be derived from formula (113) from [19] by using the correspondence rule (30):

$$Z_M^{(1)}(\alpha_1 x_1, \ldots, \alpha_M x_M) = i c_0 (-g)^{-(M+r)/2} \exp(-S_0/g) \times \int d^4 x_0 \int \frac{d\rho}{\rho^{M+5}} \exp(\nu \ln \mu\rho) \times \phi_c\left(\frac{x_1 - x_0}{\rho} \right) \ldots \phi_c\left(\frac{x_M - x_0}{\rho} \right) \int d^n u \, \delta(|\mathbf{u}| - 1) u_{\alpha_1} \ldots u_{\alpha_M}, \quad (102)$$

where $\phi_c(x)$ is the instanton solution (90) with $\rho = 1$, μ is the momentum of charge normalization,[16]

$$c_0 = \frac{1}{(2\pi)^{r/2}} \left(\frac{I_6}{4} \right)^2 J^{1/2} I_4^{(n-1)/2} \left[-\bar{D}_R(1) \bar{D}_R^{n-1}\left(\frac{1}{3} \right) \right]^{-1/2} \times \exp\left[-\frac{3}{4} r + \nu \left(-\frac{1}{2} \ln 3 + C_{\text{E}} - \frac{1}{6} \right) \right], \quad (103)$$

$$r = n + 4, \quad \nu = (n + 8)/3, \quad J = \int d^4 x \, 3 \phi_c^2(x) [\partial \phi_c(x)/\partial \rho]_{\rho = 1}^2,$$

and the expressions for S_0 and I_p are the same as in (94). The two-instanton contribution has the form

$$G_M^{(2)}(\alpha_1 x_1, \ldots, \alpha_M x_M) = -c_0^2 (-g)^{-(M+2r)/2} \exp\left(-\frac{2S_0}{g} \right) \times \int d^n u \, \delta(|\mathbf{u}| - 1) u_{\alpha_1} \ldots u_{\alpha_M} \int \frac{d\rho}{\rho^{M+5}} \exp(\nu \ln \mu\rho) \times \int d^4 x_0 \phi_c\left(\frac{x_1 - x_0}{\rho} \right) \ldots \phi_c\left(\frac{x_\mu - x_0}{\rho} \right) \times \int d^n u' \delta(|\mathbf{u}'| - 1) \int \frac{d\rho_1}{\rho_1^5} \exp(\nu \ln \mu\rho_1) \quad (104)$$

[16] For a transition from the initial to a renormalized charge in the formulas from [19], we must carry out the substitution $\ln\Lambda\rho - \ln 2 + C_E + 1/3 \longrightarrow \ln\mu\rho - (1/2)\ln 3 + C_E - 1/6$.

$$\times\int d^4x_0'\left\{\exp\left[-\frac{S_{\rm int}(\mathbf{u},\mathbf{u}',\rho,\rho_1,x_0'-x_0)}{g}\right]-1\right\}.$$

Using the relation[17]

$$\int_0^\infty\frac{d\rho_1}{\rho_1^5}\exp(\nu\ln\mu\rho_1)\int d^4R\,\delta\left(z-\frac{\rho\rho_1}{R^2+\rho_1^2}\right)=\frac{\pi^2\exp(\nu\ln\mu\rho)}{(\nu-1)(\nu-2)z^{1+\nu}},\tag{105}$$

we can transform the integral with respect to ρ_1 and x_0' in expression (104) to the form

$$\frac{\pi^2\exp(\nu\ln\mu\rho)}{(\nu-1)(\nu-2)}\times\int_0^\infty\frac{dz}{z^{1+\nu}}\exp\left(-\frac{32\pi^2(\mathbf{u}\cdot\mathbf{u}')}{g}z\right)\tag{106}$$

and obtain an integral of the type (73). Following the algorithm described in Section 6, for the asymptotic form of A_K we obtain

$$A_K=c_0\frac{2^{-M/2}}{(I_4\ln 2)^{r/2}}\frac{\pi}{4}\left(\frac{64\pi^2K}{I_4\ln 2}\right)^\nu\sigma_n\frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{1+\nu}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n+\nu}{2}\right)}\tag{107}$$

$$\times\frac{\Gamma(-\nu)(1+\cos\pi\nu)}{(\nu-1)(\nu-2)}Q\Gamma\left(K+\frac{r}{2}\right)(\ln 2)^{-K},$$

where the quantity

$$Q=\int\frac{d\rho}{\rho^{M+5}}\exp(2\nu\ln\mu\rho)\times\int d^dx_0\phi_c\left(\frac{x_1-x_0}{\rho}\right)\ldots\phi_c\left(\frac{x_M-x_0}{\rho}\right)\times\left[\int\frac{d\rho}{\rho^{M+5}}\exp(\nu\ln\mu\rho)\times\int d^dx_0\phi_c\left(\frac{x_1-x_0}{\rho}\right)\ldots\phi_c\left(\frac{x_M-x_0}{\rho}\right)\right]^{-1}\tag{108}$$

substantially depends on the external coordinates x_1, …, x_M. The expression for this quantity can be slightly simplified by passing to the momentum space and by choosing the values of external momenta $\mathbf{p}_i$ on the order of p, estimating their values at a symmetric point ($\mathbf{p}_i\cdot\mathbf{p}_j=p^2(4\delta_{ij}-1)/3$),

$$Q=\frac{\int_0^\infty dy\,y^{2M-5+2\nu}K_1^M(y)}{\int_0^\infty dy\,y^{2M-5+\nu}K_1^M(y)}\exp\left(\frac{\nu\ln\mu}{p}\right),\tag{109}$$

where $K_1(y)$ is the Macdonald function. Substituting numerical values

$$I_4=\frac{32}{3}\pi^2,\quad I_6=\frac{128}{5}\pi^2,\quad J=\frac{32}{15}\pi^2,\quad \bar D_R(1)=-578,\quad \bar D_R(1/3)=0.872\tag{110}$$

leads to

$$A_K=\frac{2^{-M/2}}{20.2}0.842^n\frac{\Gamma(-\nu)(1+\cos\pi\nu)}{(n+2)(n+5)}\times\frac{\Gamma\left(\frac{1+\nu}{2}\right)}{\Gamma\left(\frac{n+\nu}{2}\right)}Q\Gamma\left(K+\frac{r}{2}+\nu\right)(\ln 2)^{-K}\tag{111}$$

and to the numerical results given in the Introduction.

8. ASYMMETRIC INSTANTON

In this section, we consider the situation when the second instanton is definitely localized in space, but does not possess any special symmetry. In view of the absence of soft modes, we can use the algorithm developed in Section 4, according to which it is sufficient to calculate the contribution of an asymmetric instanton to the Lipatov integral (18). Following the line of reasoning in Section 7, we assume that

$$\varphi_\alpha(x)=(-g)^{-1/2}\phi_\alpha(x),\quad \phi_\alpha(x)=u_\alpha\phi_c(x),\tag{112}$$

and divide fluctuations into longitudinal and transverse:

$$\delta\varphi_\alpha(x)=\delta\varphi^L(x)u_\alpha+\delta\varphi_\alpha^T(x),\qquad\sum_\alpha\delta\varphi_\alpha^T(x)u_\alpha=0.\tag{113}$$

[17] In deriving formula (105), it is found that the main contribution to the integral appears from region $\rho_1\sim R$. In the subsequent analysis, the values of ρ_1 and R are large in view of the large value of parameter $1/g$, while the value of ρ turns out to be on the order of the reciprocal external momentum. This justifies condition (92).

Analogously to relation (20), we then obtain

$$Z_N = e^{-N}(-g_c)^{-M/2} g_c^{-N} \int_{-\infty}^{\infty} \frac{dt}{2\pi} \int D\varphi^L \times \int D\varphi_\alpha^T u_{\alpha_1} \ldots u_{\alpha_M} \phi_c(x_1) \ldots \phi_c(x_M) \times \exp\left\{ \frac{Nt^2}{2} - \frac{1}{2}(\delta\varphi^L, \hat{M}_L \delta\varphi^L) - \frac{1}{2} \sum_{\alpha=1}^{n-1} (\delta\varphi_\alpha^T, \hat{M}_T \delta\varphi_\alpha^T) \right\}, \quad (114)$$

where

$$\hat{M}_L = \hat{p}^2 + m^2 - 3\phi_c^2(x), \quad \hat{M}_T = \hat{p}^2 + m^2 - \phi_c^2(x). \quad (115)$$

The integrals with respect to longitudinal and transverse fluctuations are factorized so that zero modes for operators $\hat{M}_L$ and $\hat{M}_T$ can be singled out independently. Carrying out the orthogonal transformation of variables $\delta\varphi$, which diagonalizes $\hat{M}_L$ and $\hat{M}_T$, we perform the division in accordance with the scheme

$$\delta\varphi = \delta\varphi' + \delta\tilde{\varphi}, \quad D\varphi = D(\delta\varphi')D(\delta\tilde{\varphi}) \quad (116)$$

for $\delta\varphi^L$ and $\delta\varphi_\alpha^T$, where tildes and primes mark the subspace of zero modes and the space orthogonal to it, respectively. Performing Gaussian integration over nonzero modes, we obtain

$$Z_N = \frac{(-S\{\phi_c\})^{-(M+r)/2}}{(2\pi)^{1+r/2}} S\{\phi_c\}^{-N} \Gamma\left(N + \frac{M+r}{2}\right) \times \sqrt{-\frac{D_0}{D_L'}\left(\frac{D_0}{D_T'}\right)^{n-1}} \int D(\delta\tilde{\phi}^L)\phi_c(x_1)\ldots\phi_c(x_M) \times \int D(\delta\tilde{\phi}_\alpha^T) u_{\alpha_1}\ldots u_{\alpha_M}, \quad (117)$$

where $D_0 = \det S''\{0\}$, $D_L' = \det[\hat{M}_L]_{P'}$, and $D_T' = \det[\hat{M}_T]_{P'}$.

It is well known that the existence of zero modes is associated with the symmetry of action, $S\{\phi\} = S\{\hat{L}\phi\}$, relative to a certain continuous group of transformations defined by operator $\hat{L}$; if ϕ_c is an instanton (i.e., the solution to the equation $S'(\phi_c) = 0$), $\hat{L}\phi_c$ is also an instanton ($S'(\hat{L}\phi_c) = 0$). Using the infinitesimal form of operator $\hat{L}$ close to the unit operator, $\hat{L}_\epsilon = 1 + \epsilon\hat{T}$, we can easily see that $\hat{T}\phi_c$ is a zero mode of operator $S''\{\phi_c\}$, which is obviously connected with the generator of group $\hat{T}$. In the given case, the following groups of transformations are significant:

(a) rotations in the vector space,

$$\hat{L}^T \phi_\alpha(x) = g_{\alpha\beta}\phi_\beta(x), \quad (118)$$

where $g_{\alpha\beta}$ are the elements of an orthogonal matrix;

(b) translations

$$\hat{L}^t(x_0)\phi(x) = \phi(x + x_0); \quad (119)$$

(c) dilatation for $d = 4$, associated with the scale invariance of the massless four-dimensional theory,

$$\hat{L}^\partial(\ln\rho)\phi(x) = \rho\phi(\rho x); \quad (120)$$

(d) rotations in the coordinate space,

$$\hat{L}^r\{\theta_s\}\phi(x) = \phi(\hat{g}x), \quad (121)$$

where $\hat{g} = \hat{g}\{\theta_s\}$ is the orthogonal matrix defined by the rotational angles θ_s.

Transformation (118) is reduced to the rotation of the unit vector **u** in (112) and generates the zero mode

$$h_T(x) = \phi_c(x) \quad (122)$$

of operator $\hat{M}_T$; in expression (114), this mode is $(n-1)$-fold degenerate. The separation of this mode is performed in the conventional way [24] and corresponds to the following substitution in formula (117):

$$\int D(\delta\tilde{\phi}_T) u_{\alpha_1}\ldots u_{\alpha_M} \longrightarrow I_2^{(n-1)/2} \int d^n u \delta(|u| - 1) u_{\alpha_1}\ldots u_{\alpha_M}, \quad (123)$$

where integrals I_p are defined in (94). Infinitesimal forms of operators L^t, L^∂, and L^r can be written as

$$\hat{L}^t(\delta x_0) = 1 + \sum_i \delta x_{0,i} \frac{\partial}{\partial x_i} \equiv 1 + \sum_i \delta x_{0,i} \hat{T}_i^t,$$
$$\hat{L}^\partial(\varepsilon) = 1 + \varepsilon\left(1 + \sum_i x_i \frac{\partial}{\partial x_i}\right) \equiv 1 + \varepsilon\hat{T}^\partial,$$
$$\hat{L}^r(\delta\theta_s) = 1 + \sum_s \delta\theta_s \hat{T}_s^r,$$
$$\hat{T}_s^r \equiv \hat{T}_{(ij)}^r = x_i \frac{\partial}{\partial x_j} - x_j \frac{\partial}{\partial x_i}, \quad (124)$$

where each operator from $\hat{T}_s^r$ corresponds to rotation in one of the $d(d-1)/2$ planes (x_i, x_j) and subscript s

denotes these planes. Accordingly, the following zero modes belonging to operator $\hat{M}_L$ exist:

$$h_i^t(x) = \frac{\partial\phi_c(x)}{\partial x_i},$$
$$h^\partial(x) = \phi_c(x) + \sum_i x_i\frac{\partial\phi_c(x)}{\partial x_i}, \quad (125)$$
$$h_s^r(x) \equiv h_{(ij)}^r(x) = x_i\frac{\partial\phi_c(x)}{\partial x_j} - x_j\frac{\partial\phi_c(x)}{\partial x_i}.$$

The existence of rotational modes $h_s^r(x)$ is a specific feature of a spherically asymmetric instanton; these modes have not been considered earlier. The nontrivial moments are connected with the non-Abelian form of the group of transformations and with nonorthogonality of the basis constructed from vectors (125).

The complete group of transformations is determined by the operator

$$\hat{L}\{\theta_s, \ln\rho, x_0\} = \hat{L}^r\{\theta_s\}\hat{L}^\partial\{\ln\rho\}\hat{L}^t\{x_0\}, \quad (126)$$

so that

$$\hat{L}f(x) = \rho f(\hat{g}\rho(x + x_0)). \quad (127)$$

The infinitesimal form of this operator,

$$\hat{L}(\delta\mu_i) = 1 + \sum_i \delta\mu_i\hat{T}_i, \quad (128)$$

includes as $\hat{T}_i$ all generators $\hat{T}_i^t$, $\hat{T}^\partial$, and $\hat{T}_s^r$ introduced by relations (124), while μ_i labels the variables $x_{0,i}$, $\ln\rho$, and θ_s.

Following the algorithm used in Section 2, we introduce the expansion of unity into the integrand in expression (114) (r_L is the number of zero modes of operator $\hat{M}_L$),

$$1 = \prod_{i=1}^{r_L}\int d\lambda_i\delta\left(\lambda_i - \frac{\int d^dx\phi^4(x)f^{(i)}(x)}{\int d^dx\phi^4(x)}\right)$$
$$= (I_4)^{r_L}\prod_{i=1}^{r_L}\int d\lambda_i\delta\Big(\lambda_i\int d^dx 4\phi_c^3(x)\delta\phi(x) \quad (129)$$
$$- \int d^dx 4\phi_c^3(x)\delta\phi(x)f^{(i)}(x)\Big),$$

where we specified to a certain extent the form of functionals $f_i(\phi)$ in expression (21) by introducing the coordinate functions $f^{(i)}(x)$,[18] carried out the expansion in the vicinity of the saddle-point configuration, and chose the instanton from the condition

$$\lambda_i = \frac{\int d^dx\phi_c^4(x)f^{(i)}(x)}{\int d^dx\phi_c^4(x)}. \quad (130)$$

To prove the arbitrariness in the choice of the instanton, we represent $\phi_c(x)$ as the result of action of operator $\hat{L}^{-1}$ on a certain fixed function $\bar{\phi}_c(x)$,

$$\phi_c(x) = \hat{L}^{-1}\bar{\phi}_c(x) = \frac{1}{\rho}\bar{\phi}_c\left(\frac{\hat{g}^{-1}x}{\rho} - x_0\right). \quad (131)$$

In this case, we obtain

$$\lambda_i \equiv \frac{\int d^dx\bar{\phi}_c^4(x)\rho^{-1}\hat{L}\{\theta_s, \ln\rho, x_0\}f(x)}{\int d^dx\bar{\phi}_c^4(x)} \quad (132)$$
$$\equiv F^{(i)}\{\theta_s, \ln\rho, x_0\}.$$

Expanding $\delta\phi(x)$ in orthonormal eigenfunctions $e_j(x)$ of operator $\hat{M}_L$, we obtain

$$\delta\phi(x) = \sum_{j=1}^{r_L} C_j e_j(x) + \ldots = \sum_{j=1}^{r_L} B_j h_j(x) + \ldots; \quad (133)$$

where we have singled out the terms corresponding to the subspace of zero modes and reexpanded these terms in functions (125). Integration of $D(\delta\tilde{\phi}^L)$ in expression (117) is in fact the integration with respect to coefficients C_j:

$$\int D(\delta\tilde{\phi}_L) \longrightarrow \int\prod_{j=1}^{r_L} dC_j = \int(\det\Gamma)^{1/2}\prod_{j=1}^{r_L} dB_j, \quad (134)$$

where Γ is the Gram matrix plotted on vectors (125). Substituting expression (133) into (129) and considering that

$$\int d^dx 4\phi_c^3(x)h_j(x) = 0 \quad (135)$$

[18] In fact, the results are independent of this choice, which is manifested in the fact that functions $f^{(i)}(x)$ do not appear in the final formula (141).

for all zero modes $h_j(x)$, we obtain

$$1 = (I_4)^{r_L}\prod_i\int d\lambda_i \times\delta\left(-\sum_j B_j\int d^dx 4\phi_c^3(x) f^{(i)}(x)h_j(x) + \ldots\right). \quad (136)$$

Now we can easily prove that

$$\int d^dx 4\phi_c^3(x)h_j(x) f^{(i)}(x) = -\int d^dx\phi_c^4(x)\hat{T}_j f^{(i)}(x) \quad (137)$$

for translations and rotations, while an analogous formula with $\hat{T}_j \longrightarrow \hat{T}_j - 1$ is valid for dilatation. We perform the variation of variables $\theta_s \longrightarrow \theta_s + \delta\theta_s$, $\ln\rho \longrightarrow \ln\rho + \varepsilon$, and $x_0 \longrightarrow x_0 + \delta x_0$ in Eqs. (132) and take into account the group relation

$$\hat{L}\{\theta_s + \delta\theta_s, \ln\rho + \varepsilon, x_0 + \delta x_0\} = \hat{L}\{\delta\theta_s', \varepsilon', \delta x_0'\}\hat{L}\{\theta_s, \ln\rho, x_0\}, \quad (138)$$

where the primed and nonprimed increments do not coincide (in view of the non-Abelian nature of the group), but are connected via a linear transformation. It can easily be proved that $\varepsilon' = \varepsilon$, $\delta x_0' = \hat{g}\rho\delta x_0$, and the relation between $\delta\theta_s'$ and $\delta\theta_s$ is defined by the relation [25]

$$\hat{g}\{\theta_s + \delta\theta_s\} = \hat{g}\{\delta\theta_s'\}\hat{g}\{\theta_s\} \quad \text{or} \quad \delta\theta_s = \sum_{s'} J_{ss'}\{\theta_s\}\delta\theta_{s'}'. \quad (139)$$

Using the infinitesimal form of operator $\hat{L}\{\delta\theta_s', \varepsilon', \delta x_0'\}$, we obtain from relations (132)

$$\frac{\int d^dx\phi_c^4(x)\hat{T}_j^t f^{(i)}(x)}{\int d^dx\phi_c^4(x)} = \frac{1}{\rho}\sum_l g_{jl}\frac{\partial F^{(i)}}{\partial x_{0,l}},$$
$$\frac{\int d^dx\phi_c^4(x)(\hat{T}^\partial - 1) f^{(i)}(x)}{\int d^dx\phi_c^4(x)} = \frac{\partial F^{(i)}}{\partial\ln\rho}, \quad (140)$$
$$\frac{\int d^dx\phi_c^4(x)\hat{T}_s^r f^{(i)}(x)}{\int d^dx\phi_c^4(x)} = \sum_{s'} J_{s's}\{\theta_s\}\frac{\partial F^{(i)}}{\partial\theta_{s'}}.$$

Substituting Eqs. (137) and (140) into expression (136) and introducing the expansion of unity obtained in this way into relation (134), we obtain

$$\int\prod_j dC_j = \int\prod_i d\lambda_i\frac{\rho^d(\det\Gamma)^{1/2}}{\det\left\|\partial F^{(i)}/\partial\mu_j\right\|\det J\{\theta_s\}} = \int\prod_j d\mu_j\frac{\rho^d(\det\Gamma)^{1/2}}{\det J\{\theta_s\}} = \int\frac{\rho^d(\det\Gamma)^{1/2}}{\det J\{\theta_s\}}d^dx_0 d\ln\rho \quad (141)$$
$$\times\prod_s d\theta_s = \int\rho^d(\det\Gamma)^{1/2}d^dx_0 d\ln\rho\, d\tau_g,$$

where we assume that the quantity $\prod_s d\theta_d/\det J\{\theta_s\}$ is the definition of the invariant measure of integration of $d\tau_g$ over the group of rotations [25]. Using relation (131) and performing the substitution $x_{0,i} \longrightarrow x_{0,i}/\rho$, we obtain the sought rule of transition to collective variables,

$$\int D(\delta\tilde{\phi}_L)\phi_c(x_1)\ldots\phi_c(x_M) \longrightarrow \int(\det\Gamma)^{1/2}(d^dx_0)d\ln\rho\, d\tau_g\rho^{-M} \times\bar{\phi}_c\left(\frac{\hat{g}^{-1}(x_1 - x_0)}{\rho}\right)\ldots\bar{\phi}_c\left(\frac{\hat{g}^{-1}(x_M - x_0)}{\rho}\right), \quad (142)$$

which is valid for $d = 4$. The result for $d < 4$ can be obtained by setting $\rho = 1$ and eliminating the integration with respect to $\ln\rho$. The expression for invariant measure $d\tau_g$ depends on the method for parametrizing matrix $\hat{g}$; if we use the Euler angles θ_l^k, this expression has the form

$$d\tau_g = \prod_{k=1}^{d-1}\prod_{l=1}^{k}\sin^{l-1}\theta_l^k d\theta_l^k, \quad 0 \le \theta_1^k < 2\pi, \quad 0 \le \theta_l^k < \pi \ \ (l \ne 1). \quad (143)$$

When the Euler angles are introduced in the d-dimensional space, the rotation matrix can be written in the form [26]

$$\hat{g} = \hat{g}^{(d-1)}\ldots\hat{g}^{(2)}\hat{g}^{(1)}, \quad \hat{g}^{(k)} = \hat{g}_1(\theta_1^k)\hat{g}_2(\theta_2^k)\ldots\hat{g}_k(\theta_k^k),$$

where $\hat{g}_i(\theta) \equiv \hat{g}_{i+1,i}(\theta)$ and $\hat{g}_{ij}(\theta)$ is the matrix of rotation through angle θ in the plane (x_i, x_j). Substitution (131) is also carried out in the Gram matrix; as a result, this substitution turns out to be a function of collective variables. In fact, the dependence on ρ can be factorized, $(\det\Gamma)^{1/2} = \rho^{-4}(\det\bar{\Gamma})^{1/2}$ (for $d = 4$), and the dependence on x_0 is ruled out in view of the possibility of transition to linear combinations of rows and col-

umns in the determinant; apparently the dependence on the angles of rotation is also absent.[19]

The nonorthogonality of zero modes $h_i(x)$ is also taken into account in the transformations of determinants according to Brézin and Parisi (see notation in [4, 19]),

$$\frac{D_L'}{D_0} = \frac{\det\Gamma}{\det G}\bar{D}(1), \quad D(1) = \lim_{z\to 1}\frac{D(z)}{(1-z)^{r_L}},$$
$$\frac{D_T'}{D_0} = \frac{I_2}{I_4}\bar{D}\left(\frac{1}{3}\right), \quad \bar{D}\left(\frac{1}{3}\right) = \lim_{z\to 1/3}\frac{D(z)}{(1-3z)}; \tag{144}$$

as a result, the Gram matrix disappears in expression (117) and a matrix G with the elements

$$G_{ij} = 3\int d^dx h_i(x)\phi_c^2(x)h_j(x) \tag{145}$$

appears instead of Γ; the dependence of G on the collective variables is the same as for Γ. As a result, for $d = 4$, we obtain

$$Z_N = \frac{(-S\{\phi_c\})^{-(M+r)/2}}{(2\pi)^{1+r/2}}S\{\phi_c\}^{-N}\Gamma\left(N+\frac{M+r}{2}\right)$$
$$\times\left[-\bar{D}(1)\bar{D}^{n-1}\left(\frac{1}{3}\right)\right]^{-1/2}I_4^{(n-1)/2}\int d^nu\,\delta(|\mathbf{u}|-1)u_{\alpha_1}\ldots u_{\alpha_M}$$
$$\times\int (d^dx_0)(d\ln\rho)d\tau_g(\det\bar{G})^{1/2}\rho^{-4-M} \tag{146}$$
$$\times\bar{\phi}_c\left(\frac{\hat{g}^{-1}(x_1-x_0)}{\rho}\right)\ldots\bar{\phi}_c\left(\frac{\hat{g}^{-1}(x_M-x_0)}{\rho}\right),$$

while for $d < 4$, we must set $\rho = 1$ and eliminate integration with respect to $\ln\rho$. The normalization of determinants is carried out in the conventional way [19, 24] by separating diverging factors and compensating them by their counterterms. As a result, quantity $\bar{D}(1)$ and $\bar{D}(1/3)$ are simply replaced by $\bar{D}_R(1)$ and $\bar{D}_R(1/3)$ for $d < 4$, while for $d = 4$, the substitution

$$[-\bar{D}(1)\bar{D}^{n-1}(1/3)]^{-1/2} \longrightarrow [-\bar{D}_R(1)\bar{D}_R^{n-1}(1/3)]^{-1/2}$$
$$\times\exp(\nu\ln\mu\rho)\exp\left[-\frac{3}{4}r+\nu\left(\frac{1}{2}\ln\frac{4}{3}-\frac{\tilde{I}_4}{I_4}\right)\right] \tag{147}$$

takes place, where

$$\tilde{I}_4 = \int\frac{d^4q}{(2\pi)^4}\langle\phi_c^2\rangle_q^2\ln q \tag{148}$$

and $\langle\phi_c^2\rangle_q$ is the Fourier component of function $\phi_c^2(x)$.

To pass to the asymptotic form of A_K (see Section 4), we must carry out the substitution $\phi_c \longrightarrow \psi_c$ and $r \longrightarrow r'$ in all expressions and represent expression (146) in the form (49). Then we have for A_K formula (47) with parameters (50), where $S_1 = S\{\psi_c\}$, $b_1 - b = d(d-1)/4$, and

$$c_1 = \frac{(-S_1)^{(-(M+r'))/2}}{(2\pi)^{1+r'/2}}\left[\bar{D}_R(1)\bar{D}_R^{n-1}\left(\frac{1}{3}\right)\right]^{-1/2}I_4^{(n-1)/2}$$
$$\times\int d^nu\,\delta(|\mathbf{u}|-1)u_{\alpha_1}\ldots u_{\alpha_M}\int d^dx_0 d\tau_g(\det\bar{G})^{1/2} \tag{149}$$
$$\times\psi_c(\hat{g}^{-1}(x_1-x_0))\ldots\psi_c(\hat{g}^{-1}(x_M-x_0))$$

for $d < 4$ and

$$c_1 = \frac{(-S_1)^{-(M+r')/2}}{(2\pi)^{1+r'/2}}\left[\bar{D}_R(1)\bar{D}_R^{n-1}\left(\frac{1}{3}\right)\right]^{-1/2}$$
$$\times I_4^{(n-1)/2}\exp\left[-\frac{3}{4}r+\nu\left(\frac{1}{2}\ln\frac{4}{3}-\frac{\tilde{I}_4}{I_4}\right)\right]$$
$$\times\int d^nu\,\delta(|\mathbf{u}|-1)u_{\alpha_1}\ldots u_{\alpha_M} \tag{150}$$
$$\times\int d^dx_0 d\ln\rho\,d\tau_g(\det\bar{G})^{1/2}\rho^{-4-M}\exp(\nu\ln\mu\rho)$$
$$\times\psi_c\left(\frac{\hat{g}^{-1}(x_1-x_0)}{\rho}\right)\ldots\psi_c\left(\frac{\hat{g}^{-1}(x_M-x_0)}{\rho}\right)$$

for $d = 4$; here, $r' = r + d(d-1)/2$. All quantities appearing in this formula can be calculated if the form of instanton $\psi_c(x)$ is known.

ACKNOWLEDGMENTS

This study was financed by the Russian Foundation for Basic Research (project no. 03-02-17519).

REFERENCES

1. L. N. Lipatov, Zh. Éksp. Teor. Fiz. **72**, 411 (1977) [Sov. Phys. JETP **45**, 216 (1977)].
2. I. M. Suslov, Zh. Éksp. Teor. Fiz. **117**, 659 (2000) [JETP **90**, 571 (2000)].
3. C. Itzykson, G. Parisi, and J. B. Zuber, Phys. Rev. Lett. **38**, 306 (1977).
4. I. M. Suslov, Usp. Fiz. Nauk **168**, 503 (1998) [Phys. Usp. **41**, 441 (1998)].

[19] After the substitution $y = \hat{g}x$, generators T_i are transformed in terms of one another. For $d = 2$ and $d = 3$, the determinant of the transformation is equal to unity and the dependence of $\det\Gamma$ on $\hat{g}$ is absent; apparently, this also holds in the general case.

5. G. L. Alfimov, V. M. Eleonsky, N. E. Kulagin, *et al.*, Physica D (Amsterdam) **44**, 168 (1990); G. L. Alfimov, Mat. Model. **2**, 67 (1990); Izv. Ross. Akad. Nauk, Ser. Fiz. **60**, 12 (1996).
6. A. G. Ushveridze, Yad. Fiz. **30**, 845 (1979) [Sov. J. Nucl. Phys. **30**, 436 (1979)]; Yad. Fiz. **32**, 1446 (1980) [Sov. J. Nucl. Phys. **32**, 747 (1980)].
7. G. G. Leptukh and A. G. Ushveridze, J. Phys. A **14**, 3085 (1981).
8. E. Brézin, G. Parisi, and J. Zinn-Justin, Phys. Rev. D **16**, 408 (1977).
9. E. B. Bogomolny and V. A. Fateyev, Phys. Lett. B **71**, 93 (1977).
10. E. B. Bogomolny, Phys. Lett. B **91**, 431 (1980).
11. E. B. Bogomolny, V. A. Fateyev, and L. N. Lipatov, Sov. Sci. Rev., Sect. A **2**, 247 (1980).
12. J. Zinn-Justin, Phys. Rep. **70**, 109 (1981); J. Math. Phys. **22**, 511 (1981); **25**, 549 (1984); Nucl. Phys. B **192**, 125 (1981); **218**, 333 (1983).
13. E. B. Bogomolny, Phys. Lett. B **67**, 193 (1977).
14. G. Parisi, Phys. Lett. B **66**, 167 (1977).
15. C. M. Bender and T. T. Wu, Phys. Rev. D **7**, 1620 (1973).
16. E. B. Bogomolny and V. A. Fateyev, Phys. Lett. B **76**, 210 (1978).
17. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics*, 3rd ed. (Nauka, Moscow, 1976; Pergamon Press, Oxford, 1980).
18. I. M. Suslov, Zh. Éksp. Teor. Fiz. **111**, 1896 (1997) [JETP **84**, 1036 (1997)].
19. I. M. Suslov, Zh. Éksp. Teor. Fiz. **111**, 220 (1997) [JETP **84**, 120 (1997)].
20. J. S. Langer, Ann. Phys. (N.Y.) **41**, 108 (1967).
21. E. Brezin and G. Parisi, J. Stat. Phys. **19**, 269 (1978).
22. I. I. Balitsky, Phys. Lett. B **273**, 282 (1991).
23. V. I. Zakharov, Nucl. Phys. B **385**, 452 (1992).
24. I. M. Suslov, Zh. Éksp. Teor. Fiz. **106**, 560 (1994) [JETP **79**, 307 (1994)].
25. M. Hamermesh, *Group Theory and Its Applications to Physical Problems* (Addison-Wesley, Reading, Mass., 1962; Mir, Moscow, 1966).
26. N. Ya. Vilenkin, *Special Functions and the Theory of Group Representations* (Nauka, Moscow, 1965; Am. Math. Soc., Providence, R.I., 1968).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 99, No. 2, 2004, pp. 254–269.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 2, 2004, pp. 288–304.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Electromagnetic Form Factors of a Massive Neutrino

M. S. Dvornikov and A. I. Studenikin

Moscow State University, Vorob'evy gory, Moscow, 119992 Russia

e-mail: maxim_dvornikov@aport.ru; studenik@srd.sinp.msu.ru

Received April 14, 2004

Abstract—Electromagnetic form factors of a massive neutrino are studied in a minimally extended standard model in an arbitrary R_ξ gauge and taking into account the dependence on the masses of all interacting particles. The contribution from all Feynman diagrams to the electric, magnetic, and anapole form factors, in which the dependence of the masses of all particles as well as on gauge parameters is accounted for exactly, are obtained for the first time in explicit form. The asymptotic behavior of the magnetic form factor for large negative squares of the momentum of an external photon is analyzed and the expression for the anapole moment of a massive neutrino is derived. The results are generalized to the case of mixing between various flavors of the neutrino. Explicit expressions are obtained for the electric, magnetic, and electric dipole and anapole transitional form factors as well as for the transitional electric dipole moment. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Analysis of electromagnetic properties of neutrinos is of considerable interest in light of the recent experimental verification of the existence of a nonzero neutrino mass and mixing between different neutrino flavors [1–3]. At the present time, it remains incompletely clear whether the neutrino is a Dirac or Majorana particle. It should be noted that these two types of elementary particles possess basically different electromagnetic characteristics [4]. It is well known that a fermion with a spin of 1/2 can possess no more than four electromagnetic form factors. As a rule, these quantities are defined in terms of the electric, magnetic dipole, electric dipole, and anapole form factors (see also Section 2). However, a Majorana neutrino can exhibit its electromagnetic properties only in terms of the interaction of the anapole form factor with an external electromagnetic field.

The calculation of radiative correction to static characteristics of an elementary particle, viz., its charge determined by the values of the corresponding form factors for zero momentum transfer, is of considerable interest. In this connection, publications [5–10], in which the electromagnetic moments of neutrinos were calculated in various theoretical models, are worth mentioning. In the series of recent publications [11–13], the electric charge and the magnetic moment of a neutrino in an arbitrary R_ξ gauge were studied. It should be recalled that the corresponding form factors for zero momentum transfer are elements of the S matrix and, hence, can be measured in experiment. Thus, the electric charge and magnetic moment should be independent of the choice of the gauge. This was demonstrated in [11–13] even for a massive neutrino [13]. Analysis of radiative corrections to electromagnetic parameters of the neutrino may directly indicate which physical theory should be used beyond the range of the standard model and provide important information on the parameters and structure of the proposed model of interaction between elementary particles. For example, for particles described in the framework of the theories with broken CP invariance, an electric dipole moment inevitably appears.

For nonzero momentum transfer, the electromagnetic form factors are not invariants of the gauge transformation group and, hence, are not measurable quantities. However, in analyzing some processes (e.g., calculating higher order corrections), the values of the electromagnetic form factors of neutrinos with a nonzero momentum transfer must be taken into account. One of such processes corresponding to radiative corrections to scattering of a neutrino by a lepton is shown in Fig. 1.

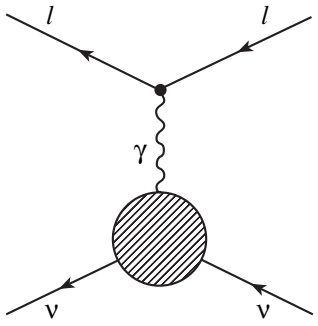


Fig. 1. A diagram illustrating the contribution to the elastic scattering of a neutrino by a lepton. The hatched circle schematically represents the neutrino electromagnetic vertex function.

The electric and magnetic form factors of elementary particles were analyzed in [6, 14] in the content of various gauge theories The case of a zero-mass neutrino was studied and not only static electromagnetic parameters, but also the asymptotic behavior of the magnetic form factors for high negative squares of the external photon momentum were considered. It is well known that the magnetic moment of a neutrino in the minimally extended standard model is proportional to the particle mass. In a number of theoretical models (see, for example, [15]), the magnitude of the magnetic moment depends on the neutrino mass only slightly and is completely determined by the mass of the heavy particle in the polarization loop.

One of electromagnetic properties of an elementary particle is its charge radius, which has been studied in many publications. The expression for the charge radius of a zero-mass neutrino in the standard model was derived in [16]. It was found that the charge radius for a zero-mass particle is a diverging quantity; moreover, it depends on the choice of gauge. In this connection, the concept of electroweak radius was introduced in [16, 17]. This quantity is determined by radiative corrections to the process of scattering of a lepton by a neutrino. The electroweak radius of a neutrino is connected with the effective Weinberg angle. Obviously, a quantity defined in this way is finite and gauge invariant. Calculation of the electroweak radius of a zero-mass neutrino in the framework of the standard model in an arbitrary R_ξ gauge was carried out in [17].

The neutrino anapole moment was considered by many authors. Among the corresponding publications, work [18], in which the anapole moment is shown to be gauge-dependent in the frame of the standard model and, hence, unobservable, is worth mentioning. In a series of publications [19, 20] using the dispersion relation method, the expression for the neutrino anapole moment was derived in the 't Hoft–Feynman gauge and the dependence of the anapole form factor on the square of the external photon momentum was studied. We must also mention the article [21], in which the expression for the anapole moment of a zero-mass neutrino was derived on the basis of the electroweak radius. As a matter of fact, a zero-mass particle is characterized by a certain relation between these quantities (see also Section 5). Thus, knowledge of one of the electromagnetic parameters (in the present case, the electroweak radius) makes it possible to easily reconstruct another characteristic and to derive an expression for the neutrino electroweak anapole moment.

In this study, we analyze the electromagnetic vertex form factors of a massive Dirac neutrino in the framework of the minimally extended standard model supplemented with a SU(2)-singlet right-handed neutrino. All calculations are made in an arbitrary R_ξ gauge, while enables us to study the dependence of the results on gauge parameters for both W and Z bosons. It should be noted that the masses of the neutrino and the charged lepton were never fixed; consequently, our analysis makes it possible to consider the limit of not only light, but also heavy neutrinos. In Section 2, the Feynman amplitudes are determined for all contributions to the electromagnetic vertex function of a massive neutrino. The expressions for the contributions of self-energy γ–Z diagrams are calculated in explicit form containing a single definite integral with respect to the Feynman parameter. The ultraviolet divergences emerging in the analysis of the neutrino electromagnetic vertex function are also treated in Section 2. In Sections 3 and 4, the contributions of all Feynman diagrams to the charge and magnetic form factors of a massive neutrino are studied. The asymptotic behavior of the magnetic form factor is analyzed for large negative squares of the external photon momentum. The anapole form factor and the anapole moment of a massive neutrino are considered in Section 5. The results of this study can be generalized to the case of mixing between different flavors of the neutrino. In particular, transitional electromagnetic form factors are studied in Section 6 in the framework of the minimally extended standard model permitting mixing between different flavors of charged leptons and neutrino. Explicit expressions are derived for the charge, magnetic, and electric dipole and transient anapole form factors. The cases of mass-degenerate and nondegenerate neutrino states are studied. Moreover, the expression for the transitional electric dipole moment is also derived in Section 6.

2. NEUTRINO VERTEX FUNCTION

The matrix element of electromagnetic current averaged over the neutrino states can be written in the form

$$\langle \nu(p')\left|J_\mu^{\mathrm{EM}}\right|\nu(p)\rangle = \bar{u}(p')\Lambda_\mu(q)u(p), \qquad (2.1)$$

where the most general expression for the electromagnetic vertex function $\Lambda_\mu(q)$ is

$$\begin{aligned}\Lambda_\mu(q) &= f_Q(q^2)\gamma_\mu + f_M(q^2)i\sigma_{\mu\nu}q^\nu \\ &- f_E(q^2)\sigma_{\mu\nu}q^\nu\gamma_5 + f_A(q^2)(q^2\gamma_\mu - q_\mu \not{q})\gamma_5.\end{aligned} \qquad (2.2)$$

Here, $f_Q(q^2)$, $f_M(q^2)$, $f_E(q^2)$, and $f_A(q^2)$ are the electric, magnetic and electric dipole, and anapole form factors of the neutrino; $q_\mu = p'_\mu - p_\mu$, $\sigma_{\mu\nu} = (i/2)[\gamma_\mu, \gamma_\nu]$, and $\gamma_5 = -i\gamma^0\gamma^1\gamma^2\gamma^3$. The values of these form factors for $q^2 = 0$ determine the static electromagnetic properties of neutrinos. In the case of a Dirac neutrino, which will be considered here, the assumptions concerning the CP invariance and the hermiticity of the electromagnetic current operator J_μ^{EM} lead to zero value of the dipole electric form factor. At zero momentum transfer, only $f_Q(0)$ and $f_M(0)$, which are known as the electric charge and magnetic moment, make a contribution to Hamiltonian $H_{\mathrm{int}} \sim J_\mu^{\mathrm{EM}}A^\mu$, which describes the interaction of the neutrino with the external electromagnetic field A^μ.

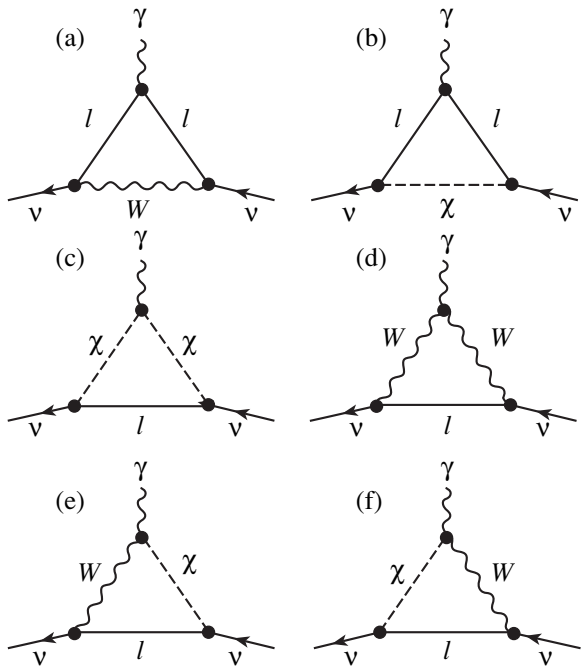


Fig. 2. Proper vertices diagrams.

The expressions for the electromagnetic function of a massive and a zero-mass neutrino differ radically. If we consider a massless elementary particle, relation (2.2) implies that the matrix element of the electromagnetic current can be written using only one form factor (see, for example, [21]),

$$\bar{u}(p')\Lambda_\mu(q)u(p) = f_D(q^2)\bar{u}(p')\gamma_\mu(1+\gamma_5)u(p).$$

It follows hence that the electric and anapole form factors are connected with function $f_D(q^2)$ via obvious relations:

$$f_Q(q^2) = f_D(q^2), \quad f_A(q^2) = f_D(q^2)/q^2.$$

However, in the case of a massive elementary particle, a simple relation connecting the electric and anapole form factor does not exist since we cannot disregard matrix terms of the form $q_\mu \not{q}\gamma_5$ in a term proportional to the anapole form factor. Moreover, direct calculation of the neutrino electromagnetic form factors revealed that, in addition to the well-known form factors given in relation (2.2), each Feynman diagram makes a nonzero contribution to the additional term proportional to matrix $\gamma_\mu\gamma_5$. These contributions differ from zero even for $q^2 = 0$. It was found in our earlier publication [13] that the sum of these contributions from all Feynman diagrams to the additional "form factor" is zero for $q^2 = 0$. The equality to zero of the "form factor" in question for $q^2 \neq 0$ in a special gauge was also proved in [13].

We will consider the calculation of one-loop Feynman diagrams for the vertex electromagnetic function of a massive neutrino in the framework of the minimally extended standard model with an SU(2)-singlet right-handed neutrino in an arbitrary R_ξ gauge. These diagram can be divided into two types: triangular (Fig. 2) and γ–Z diagrams (Fig. 4). Using Feynman's rules formulated in [22], we can determine the contributions to the neutrino vertex function $\Lambda_\mu(q)$. Applying dimensional regularization in the corresponding Feynman integrals, we find that the contributions from the proper vertices diagrams (Fig. 2) can be written in the form

$$\Lambda_\mu^{(1)} = i\frac{eg^2}{2}\int\frac{d^Nk}{(2\pi)^N}\left[g^{\kappa\lambda} - (1-\alpha)\frac{k^\kappa k^\lambda}{k^2-\alpha M_W^2}\right]\frac{\gamma_\kappa^L(\not{p}' - \not{k} + m_l)\gamma_\mu(\not{p} - \not{k} + m_l)\gamma_\lambda^L}{[(p'-k)^2 - m_l^2][(p-k)^2 - m_l^2][k^2 - M_W^2]}, \tag{2.3}$$

$$\Lambda_\mu^{(2)} = i\frac{eg^2}{2M_W^2}\int\frac{d^Nk}{(2\pi)^N}\frac{(m_\nu P_L - m_l P_R)(\not{p}' - \not{k} + m_l)\gamma_\mu(\not{p} - \not{k} + m_l)(m_l P_L - m_\nu P_R)}{[(p'-k)^2 - m_l^2][(p-k)^2 - m_l^2][k^2 - \alpha M_W^2]}, \tag{2.4}$$

$$\Lambda_\mu^{(3)} = i\frac{eg^2}{2M_W^2}\int\frac{d^Nk}{(2\pi)^N}(2k-p-p')_\mu\frac{(m_\nu P_L - m_l P_R)(\not{k} + m_l)(m_l P_L - m_\nu P_R)}{[(p'-k)^2 - \alpha M_W^2][(p-k)^2 - \alpha M_W^2][k^2 - m_l^2]}, \tag{2.5}$$

$$\begin{aligned}\Lambda_\mu^{(4)} = i\frac{eg^2}{2}\int\frac{d^Nk}{(2\pi)^N}\gamma_k^L(\not{k} + m_l)\gamma_\lambda^L\left[\delta_\beta^\kappa - (1-\alpha)\frac{(p'-k)^\kappa(p'-k)_\beta}{(p'-k)^2 - \alpha M_W^2}\right]\left[\delta_\gamma^\lambda - (1-\alpha)\frac{(p-k)^\lambda(p-k)_\gamma}{(p-k)^2 - \alpha M_W^2}\right]\\ \times\frac{\delta_\mu^\beta(2p'-p-k)^\gamma + g^{\beta\gamma}(2k-p-p')_\mu + \delta_\mu^\gamma(2p-p'-k)^\beta}{[(p'-k)^2 - M_W^2][(p-k)^2 - M_W^2][k^2 - m_l^2]},\end{aligned} \tag{2.6}$$

$$\Lambda_\mu^{(5)+(6)} = i\frac{eg^2}{2}\int\frac{d^Nk}{(2\pi)^N}\left\{\frac{\gamma_\beta^L(\hat{k}-m_l)(m_lP_L-m_\nu P_R)}{[(p'-k)^2-M_W^2][(p-k)^2-\alpha M_W^2][k^2-m_l^2]}\left[\delta_\mu^\beta-(1-\alpha)\frac{(p'-k)^\beta(p'-k)_\mu}{(p'-k)^2-\alpha M_W^2}\right]\right.$$
$$\left.-\frac{(m_\nu P_L-m_lP_R)(\hat{k}-m_l)\gamma_\beta^L}{[(p'-k)^2-\alpha M_W^2][(p-k)^2-M_W^2][k^2-m_l^2]}\left[\delta_\mu^2-(1-\alpha)\frac{(p-k)^\beta(p-k)_\mu}{(p-k)^2-\alpha M_W^2}\right]\right\}. \quad (2.7)$$

Here, m_ν, M_W, and m_l are the masses of the neutrino, the W boson, and the charged lepton (which is the lower component of the isodoublet relative to the neutrino); e is the proton charge; g is the coupling constant in the standard model; θ_W is the Weinberg angle; $\alpha = 1/\xi$ is the gauge parameter for the W boson; and $P_{L,R} = (1 \pm \gamma_5)/2$ are the projection operators.

The contributions from the γ–Z diagrams (Fig. 4) to the vertex function $\Lambda_\mu(q)$ are shown in Fig. 3 and are given by the following expressions:

$$\Lambda_\mu^{(j)}(q) = \frac{g}{2\cos\theta_W}\Pi_{\mu\nu}^{(j)}(q)\frac{1}{q^2-M_Z^2}\times\left\{g^{\nu a}-(1-\alpha_Z)\frac{q^\nu q^\alpha}{q^2-\alpha_ZM_Z^2}\right\}\gamma_\alpha^L, \quad (2.8)$$
$$j = 7, \ldots, 14,$$

where

$$\Pi_{\mu\nu}^{(7)}(q) = -ieg\cos\theta_W\times\int\frac{d^Nk}{(2\pi)^N}\frac{1}{[(k-q)^2-M_W^2][k^2-M_W^2]}$$
$$\times\left[g_{\gamma\alpha}-(1-\alpha)\frac{(k-q)_\gamma(k-q)_\alpha}{(k-q)^2-\alpha M_W^2}\right]\times\left[g_{\beta\lambda}-(1-\alpha)\frac{k_\beta k_\lambda}{k^2-\alpha M_W^2}\right] \quad (2.9)$$
$$\times[(k+q)^\gamma\delta_\mu^\beta+(q-2k)_\mu g^{\beta\gamma}+(k-2q)^\beta\delta_\mu^\gamma]\times[(k+q)^\alpha\delta_\nu^\lambda+(q-2k)_\nu g^{\alpha\lambda}+(k-2q)^\lambda\delta_\nu^\alpha],$$

$$\Pi_{\mu\nu}^{(8)}(q) = -2ieg\frac{\sin^2\theta_W}{\cos\theta_W}M_W^2\times\int\frac{d^Nk}{(2\pi)^N}\frac{1}{[(k-q)^2-\alpha M_W^2][k^2-M_W^2]}\times\left[g_{\mu\nu}-(1-\alpha)\frac{k_\mu k_\nu}{k^2-\alpha M_W^2}\right], \quad (2.10)$$

$$\Pi_{\mu\nu}^{(9)}(q) = ieg\frac{\cos^2\theta_W-\sin^2\theta_W}{\cos\theta_W}\times\int\frac{d^Nk}{(2\pi)^N}\frac{g_{\mu\nu}}{k^2-\alpha M_W^2}, \quad (2.11)$$

$$\Pi_{\mu\nu}^{(10)}(q) = -ieg\cos\theta_W\times\int\frac{d^Nk}{(2\pi)^N}\frac{\delta_\mu^\alpha\delta_\nu^\beta+\delta_\mu^\beta\delta_\nu^\alpha-2g^{\alpha\beta}g_{\mu\nu}}{k^2-M_W^2}\times\left[g_{\alpha\beta}-(1-\alpha)\frac{k_\alpha k_\beta}{k^2-\alpha M_W^2}\right], \quad (2.12)$$

$$\Pi_{\mu\nu}^{(11)+(12)}(q) = 2ieg\cos\theta_W\times\int\frac{d^Nk}{(2\pi)^N}\frac{k_\mu(k-q)_\nu}{[(k-q)^2-\alpha M_W^2][k^2-\alpha M_W^2]}, \quad (2.13)$$

$$\Pi_{\mu\nu}^{(13)}(q) = ieg\frac{\sin^2\theta_W-\cos^2\theta_W}{2\cos\theta_W}\int\frac{d^Nk}{(2\pi)^N}(2k-q)_\mu\times(2k-q)_\nu\frac{1}{[(k-q)^2-\alpha M_W^2][k^2-\alpha M_W^2]}, \quad (2.14)$$

$$\Pi_{\mu\nu}^{(14)}(q) = \frac{ieg}{2\cos\theta_W}\sum_f Q_f\int\frac{d^Nk}{(2\pi)^N}\times\frac{1}{[(k-q)^2-m_f^2][k^2-m_f^2]}\mathrm{Tr}\left[\gamma_\mu(\hat{k}+m_f)\gamma_\nu\times\left(\pm\frac{1}{2}-2Q_f\sin^2\theta_W\pm\frac{1}{2}\gamma_5\right)(\hat{k}-\hat{q}+m_f)\right]. \quad (2.15)$$

Here, M_Z and α_Z denote the mass and the gauge parameter of the Z boson. The minus and plus signs in expression (2.15) correspond to the "upper" (u, c, and t quarks) and "lower" (electron, muon, τ lepton as well as d, s, and b quarks) components of the isodoublet and m_f and Q_f are the mass and the electric charge (in units of e) of a fermion in the loop.

It will be convenient in the subsequent analysis to expand each contribution from the γ–Z diagrams for an

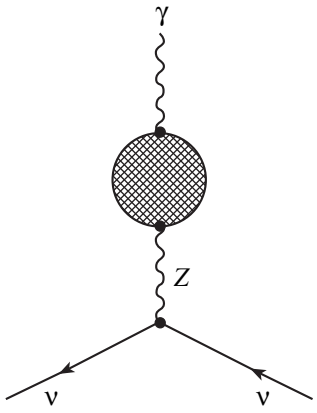


Fig. 3. Contributions of the γ–*Z* diagrams to the neutrino electromagnetic vertex function. Hatched circle schematically represents the function $\Pi^{(j)}_{\mu\nu}(q)$.

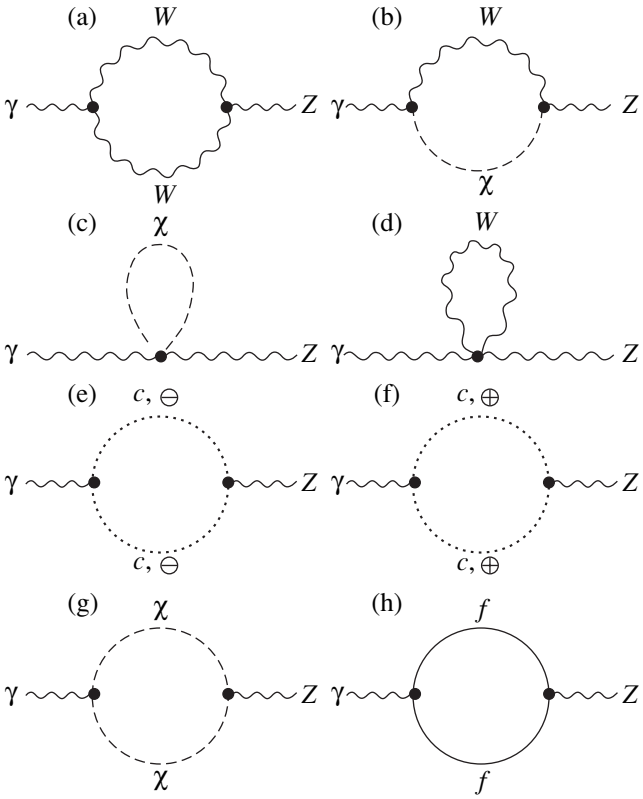


Fig. 4. γ–*Z* diagrams: *f* denotes an electron, muon, τ lepton as well as the *u*, *c*, *t*, *d*, *s*, and *b* quarks.

arbitrary q^2 and to explicitly separate the transverse part:

$$\Pi^{(j)}_{\mu\nu}(q) = A^{(j)}(\alpha, q^2)\left(g_{\mu\nu} - \frac{q_\mu q_\nu}{q^2}\right) + B^{(j)}(\alpha, q^2)g_{\mu\nu}, \quad j = 7, \ldots, 14. \tag{2.16}$$

Using expressions (2.9)–(2.15) for the contributions from the γ–*Z* diagrams in the form of Feynman integrals as well as formula (2.16), we can write the functions $A^{(j)}(a, q^2)$ and $B^{(j)}(a, q^2)$, where $j = 7, \ldots, 14$, in explicit form:

$$\begin{aligned} A^{(7)}(\alpha, q^2) &= 2M_W^2\cos^3\theta_W\sin\theta_W M_Z^2\tilde{G}_F\tau \\ &\times\left[\omega\left(-\frac{14}{3}+\alpha\right)+\frac{1}{6}+\frac{\alpha}{2}\right. \\ &-2\tau\int_0^1 dx(1-x^2)^2\{\ln(1-\zeta-x(1-\alpha))-\ln(1-\zeta)\} \\ &+2\int_0^1 dx(5x^2-5x-1)\ln(1-\zeta) \\ &-2\int_0^1 dx(4x^2-3)\{(1-\zeta-x(1-\alpha)) \\ &\times\ln(1-\zeta-x(1-\alpha))-(1-\zeta)\ln(1-\zeta)\} \\ &+\frac{\tau}{2}\int_0^1 dx\{2(1-\zeta-x(1-\alpha))\ln(1-\zeta-x(1-\alpha)) \\ &\left.-(1-\zeta)\ln(1-\zeta)-(\alpha-\zeta)\ln(\alpha-\zeta)\}\right], \end{aligned} \tag{2.17}$$

$$\begin{aligned} A^{(8)}(\alpha, q^2) &= -4M_W^2\cos\theta_W\sin^3\theta_W M_Z^2\tilde{G}_F\tau \\ &\times\int_0^1 dx x^2\{\ln(1-\zeta-x(1-\alpha))-\ln(\alpha-\zeta)\}, \end{aligned} \tag{2.18}$$

$$A^{(9)}(\alpha, q^2) = 0, \tag{2.19}$$

$$A^{(10)}(\alpha, q^2) = 0, \tag{2.20}$$

$$\begin{aligned} A^{(11)+(12)}(\alpha, q^2) &= 2M_W^2\cos^3\theta_W\sin\theta_W M_Z^2\tilde{G}_F\tau \\ &\times\left[\frac{\omega}{3}+2\int_0^1 dx x(1-x)\ln(\alpha-\zeta)\right], \end{aligned} \tag{2.21}$$

$$\begin{aligned} A^{(13)}(\alpha, q^2) &= M_W^2(\sin^2\theta_W-\cos^2\theta_W)\cos\theta_W\sin\theta_W \\ &\times M_Z^2\tilde{G}_F\tau\left[-\frac{\omega}{3}-\int_0^1 dx(2x-1)^2\ln(\alpha-\zeta)\right], \end{aligned} \tag{2.22}$$

$$A^{(14)}(\alpha, q^2) = 8M_W^2\cos\theta_W\sin\theta_W M_Z^2\tilde{G}_F\tau$$

$$\times\left[\frac{\omega}{6}\left(-3-\frac{28}{3}\sin^2\theta_W\right)\right.$$
$$\left.+\sum_f Q_f\left(\pm\frac{1}{2}-2Q_f\sin^2\theta_W\right)\left\{\frac{1}{6}\ln\left(\frac{m_f}{M_W}\right)^2\right.\right.$$
$$\left.\left.+\int_0^1 dx x(1-x)\ln\left(1-\left(\frac{M_W}{m_f}\right)^2\zeta\right)\right\}\right], \quad (2.23)$$

$$B^{(7)}(\alpha,q^2) = 2M_W^2\cos^3\theta_W\sin\theta_W M_Z^2\tilde{G}_F$$
$$\times\left[\omega\left(\frac{\tau}{2}-\frac{12+3\alpha(1+\alpha)}{2}\right)\right.$$
$$+\frac{3}{4}(2+\alpha(1+\alpha))-\frac{\tau}{24}(25+3\alpha)$$
$$-3\tau\int_0^1 dx(2x-1)^2\ln(1-\zeta)-9\int_0^1 dx(1-\zeta)\ln(1-\zeta) \quad (2.24)$$
$$-3\tau\int_0^1 dx x^2\{(1-\zeta-x(1-\alpha))\ln(1-\zeta-x(1-\alpha))$$
$$-(1-\zeta)\ln(1-\zeta)\}-\frac{9}{2}\int_0^1 dx\{(1-\zeta-x(1-\alpha))^2$$
$$\left.\times\ln(1-\zeta-x(1-\alpha))-(1-\zeta)^2\ln(1-\zeta)\}\right],$$

$$B^{(8)}(\alpha,q^2) = 2M_W^2\cos\theta_W\sin^3\theta_W M_Z^2\tilde{G}_F$$
$$\times\left[-\omega\frac{3+\alpha}{2}-\frac{1-\alpha}{2}-2\int_0^1 dx\ln(1-\zeta-x(1-\alpha))\right.$$
$$+\int_0^1 dx\{(1-\zeta-x(1-\alpha))\ln(1-\zeta-x(1-\alpha)) \quad (2.25)$$
$$-(\alpha-\zeta)\ln(\alpha-\zeta)\}$$
$$\left.+2\tau\int_0^1 dx x^2\{\ln(1-\zeta-x(1-\alpha))-\ln(\alpha-\zeta)\}\right],$$

$$B^{(9)}(\alpha,q^2) = 2M_W^2(\cos^2\theta_W-\sin^2\theta_W)\cos\theta_W$$
$$\times\sin\theta_W M_Z^2\tilde{G}_F[\alpha(\omega-1)+\alpha\ln\alpha], \quad (2.26)$$

$$B^{(10)}(\alpha,q^2) = 6M_W^2\cos^3\theta_W\sin\theta_W M_Z^2\tilde{G}_F$$
$$\times\left[\omega\frac{3+\alpha^2}{2}-\frac{1}{4}-\frac{5\alpha^2}{12}+\frac{\alpha^2\ln\alpha}{2}\right], \quad (2.27)$$

$$B^{(11)+(12)}(\alpha,q^2) = 2M_W^2\cos^3\theta_W\sin\theta_W M_Z^2\tilde{G}_F$$
$$\times\left[\omega\left(\alpha-\frac{\tau}{2}\right)-\alpha+\frac{\tau}{6}+\int_0^1 dx(\alpha-\zeta)\ln(\alpha-\zeta)\right.$$
$$\left.-2\tau\int_0^1 dx x(1-x)\ln(\alpha-\zeta)\right], \quad (2.28)$$

$$B^{(13)}(\alpha,q^2) = 2M_W^2(\sin^2\theta_W-\cos^2\theta_W)\cos\theta_W\sin\theta_W$$
$$\times M_Z^2\tilde{G}_F\left[\alpha(\omega-1)+\frac{\tau}{6}+\int_0^1 dx(\alpha-\zeta)\ln(\alpha-\zeta)\right.$$
$$\left.+\frac{\tau}{2}\int_0^1 dx(2x-1)^2\ln(\alpha-\zeta)\right], \quad (2.29)$$

$$B^{(14)}(\alpha,q^2) = 0, \quad (2.30)$$

where

$$\tilde{G}_F=\frac{G_F}{4\pi^2\sqrt{2}},\quad \omega=-\frac{1}{\varepsilon}-\ln(4\pi^2)+\mathbb{C}-\ln\frac{\lambda^2}{M_W^2},$$

G_F is the Fermi constant, $\zeta=\tau x(1-x)$, $\tau=q^2/M_W^2$, ε and λ are the constants introduced during dimensional regularization, and $\mathbb{C}$ is the Euler constant.

In deriving relations (2.17)–(2.30), we used the properties of the algebra of γ matrices in the N-dimensional space and the expressions for the characteristic Feynman integrals given in [22, 23]. Note that $\varepsilon = 2 - N/2 > 0$, where N is the dimension of space. When regularization is removed, $\varepsilon \longrightarrow 0$.

Let us now consider ultraviolet divergences emerging in the calculation of the electromagnetic vertex function. The sum of the contributions of the diverging parts of proper vertices diagrams (2.3)–(2.7) to the electromagnetic vertex function of a massive neutrino has the form

$$\Lambda_\mu^{(\text{div.prop.vert.})}(q) = -\frac{eG_F}{4\pi^2\sqrt{2}}M_W^2\omega\frac{3+\alpha}{2}\gamma_\mu^L. \quad (2.31)$$

Note that this expression is independent of the external photon momentum q_μ.

In the subsequent analysis of diverging contributions from the γ–Z diagrams (2.9)–(2.15), it is convenient to use relations (2.8) and (2.16). Using these formulas, we obtain the following expression for the sum

of the contributions from the γ–Z diagrams to the electromagnetic vertex function of a massive neutrino:

$$\Lambda_\mu^{(\gamma-Z)}(q) = \frac{g}{4\cos\theta_W}\Bigg[\frac{A(\alpha, q^2) + B(\alpha, q^2)}{q^2 - M_Z^2}\gamma_\mu + \frac{1}{q^2 - M_Z^2}\Bigg\{\frac{A(\alpha, q^2)}{q^2} + (1-\alpha_Z)\frac{B(\alpha, q^2)}{q^2 - \alpha_Z M_Z^2}\Bigg\} \times (q^2\gamma_\mu - q_\mu \not{q})\gamma_5 + \frac{\alpha_Z B(\alpha, q^2)}{q^2 - \alpha_Z M_Z^2}\gamma_\mu\gamma_5\Bigg]. \quad (2.32)$$

The diverging parts of functions $A(\alpha, q^2)$ and $B(\alpha, q^2)$ have the form

$$A^{\mathrm{div}}(\alpha, q^2) = 2M_W^2\cos\theta_W\sin\theta_W M_Z^2\tilde{G}_F\tau\omega \times\Bigg\{\Big(\alpha - \frac{37}{6}\Big)\cos^2\theta_W - \frac{151}{18}\sin^2\theta_W\Bigg\}, \quad (2.33)$$

$$B^{\mathrm{div}}(\alpha, q^2) = -2M_W^2\cos\theta_W\sin\theta_W M_Z^2\tilde{G}_F\frac{3+\alpha}{2}. \quad (2.34)$$

Formulas (2.31)–(2.34) imply that all the form factors except the magnetic one contain divergences and depend on the choice of the gauge (both on α and on α_Z). In spite of this, we can choose the gauge parameters in such a way that the complete expression for $\Lambda_\mu(q)$ including the contributions from the triangular (Fig. 2) and γ–Z diagrams (Fig. 4) does not contain ultraviolet divergences. Indeed, fixing the gauge parameters as

$$\alpha = \frac{1}{9}(138 + 151\tan^2\theta_W), \quad \alpha_Z = +\infty,$$

we find that all the terms in $\Lambda_\mu(q)$ containing the pole $1/\varepsilon$ mutually cancel out. Thus, in the given gauge, the electromagnetic vertex function of a massive neutrino is finite in the one-loop approximation for an arbitrary external photon momentum q_μ.

An analogous statement can be formulated for the case of the electron electromagnetic vertex function in quantum electrodynamics. The expression for the electron vertex function in the one-loop approximation is given in [23] in an arbitrary gauge. Using formula (24'), p.358, in [23], we find that all form factors in the vertex function are finite for $d_l = 3$, where d_l is the photon gauge parameter.

3. ELECTRIC FORM FACTOR OF A NEUTRINO

In this section, we consider the electric form factor of a massive neutrino. Using the results obtained in the preceding section for various contributions to the neutrino vertex function $\Lambda_\mu(q)$, we single out in formulas (2.3)–(2.15) the coefficients proportional to matrix γ_μ, which are the corresponding contributions to the electric form factor $f_Q(q^2)$ in accordance with the expansion given in relation (2.2).

First of all, we consider the contributions of one-loop proper vertices diagrams (Fig. 2) to the neutrino electric from factor. Using the familiar identity

$$\bar{u}(p)(p'_\mu + p_\mu)u(p) = \bar{u}(p')(2m_\nu\gamma_\mu - i\sigma_{\mu\nu}q^\nu)u(p),$$

and integrating over the momenta of virtual particles (the details of this procedure for dimensional regularization are given in [23]), we obtain exact expressions for the contributions of the diagrams considered here to the electric form factor of a massive neutrino in terms of definite integrals:

$$f_Q^{(\mathrm{prop.vert.})}(q^2) = \frac{eG_F}{4\pi^2\sqrt{2}}M_W^2\sum_{i=1}^{6}\bar{f}_Q^{(i)}(q^2).$$

Here,

$$\bar{f}_Q^{(1)}(q^2) = \omega\frac{\alpha}{2} + 1 + \frac{1-\alpha}{12} + \int_0^1 dz\int_0^z dy\ln\mathfrak{D}_1 - \int_0^1 dz\int_0^z dy[a + b(1-z)^2 + \tau(1 - z + y(z-y))]\frac{1}{\mathfrak{D}_1} + \frac{1}{2}\int_0^1 dz\int_0^z dy[bz^2(1 + b(1-z)^2) + a\tau y(z-y) + b\tau(2zy(z-y)(1-z) + 5y(z-y) - z^2(1-z)) + \tau^2 y(z-y)(1 - z + yz - y^2)\Big[\frac{1}{\mathfrak{D}_1(\alpha)} - \frac{1}{\mathfrak{D}_1}\Big] \quad (3.1)$$

$$-\frac{1}{2}\int_0^1 dz\int_0^z dy[a + b + 6bz(1-z) + \tau(1 - 3z + 6y(z-y))] \times[\ln\mathfrak{D}_1(\alpha) + \ln\mathfrak{D}_1],$$

$$\bar{f}_Q^{(2)}(q^2) = \frac{a+b}{2}\Bigg(\frac{\omega}{2} + \frac{1}{2} + \int_0^1 dz\int_0^z dy\ln\mathfrak{D}_1(\alpha)\Bigg) - \frac{1}{2}\int_0^1 dz\int_0^z dy(a^2 + abz^2 + b^2z^2 - 4abz + ab + (a+b)\tau y(z-y))\frac{1}{\mathfrak{D}_1(\alpha)}, \quad (3.2)$$

$$\bar{f}_Q^{(3)}(q^2) = \frac{a+b}{2}\left(-\frac{\omega}{2} - \int_0^1 dz \int_0^z dy \ln \mathfrak{D}_2(\alpha)\right)$$
$$+ b\int_0^1 dz\int_0^z dy(3az - az^2 - 2a + bz(1-z))\frac{1}{\mathfrak{D}_2(\alpha)}, \quad (3.3)$$

$$\bar{f}_Q^{(4)}(q^2) = -\omega\frac{3}{4}(1+\alpha) - 1 - 3\int_0^1 dz\int_0^z dy \ln \mathfrak{D}_2$$
$$+ \int_0^1 dz\int_0^z dy(3bz(1-z) - \tau(z - y(z-y)))\frac{1}{\mathfrak{D}_2}$$
$$-\frac{9}{2}\int_0^1 dz\int_0^z dy[(\mathfrak{D}_2(\alpha) + y(1-\alpha))$$
$$\times \ln(\mathfrak{D}_2(\alpha) + y(1-\alpha)) - \mathfrak{D}_2 \ln \mathfrak{D}_2]$$
$$-\int_0^1 dz\int_0^z dy[2b^2(1-z)^2(z(1-z) - y)$$
$$- b\tau(y(z-y)(5z - 3z^2 - 3y) + z(1-z)^2 - y(2 - y - y^2))$$
$$-\tau^2 y(z-y)(1 - z + yz + y + y^2)]$$
$$\times\left[\frac{1}{\mathfrak{D}_2(\alpha) + y(1-\alpha)} - \frac{1}{\mathfrak{D}_2}\right] \quad (3.4)$$
$$+\frac{1}{2}\int_0^1 dz\int_0^z dy[3b(1-z^2) + \tau(4 - 6(z-y) + 11y(z-y))]$$
$$\times[\ln(\mathfrak{D}_2(\alpha) + y(1-\alpha)) - \ln \mathfrak{D}_2]$$
$$-\frac{b\tau}{2}\int_0^1 dz\int_0^z dy(bz(1 - 3z + z^2 + z^3)$$
$$-\tau y(z-y)(z + z^2 - 2y))$$
$$\times\left[\frac{1}{\mathfrak{D}_2} + \frac{1}{\mathfrak{D}_2(\alpha)} - \frac{2}{\mathfrak{D}_2(\alpha) + y(1-\alpha)}\right]$$
$$+\frac{\tau}{4}\int_0^1 dz\int_0^z dy(b(9 - 13z + 4z^2) - 2\tau y(z-y))$$
$$\times[\ln \mathfrak{D}_2 + \ln \mathfrak{D}_2(\alpha) - 2\ln(\mathfrak{D}_2(\alpha) + y(1-\alpha))]$$
$$+\frac{3\tau}{4}\int_0^1 dz\int_0^z dy[\mathfrak{D}_2 \ln \mathfrak{D}_2 + \mathfrak{D}_2(\alpha)\ln \mathfrak{D}_2(\alpha)$$
$$-2(\mathfrak{D}_2(\alpha) + y(1-\alpha))\ln(\mathfrak{D}_2(\alpha) + y(1-\alpha))],$$

$$\bar{f}_Q^{(5)+(6)}(q^2) = \int_0^1 dz\int_0^z dy(a - bz)\frac{1}{\mathfrak{D}_2(\alpha) + y(1-\alpha)}$$
$$-b\int_0^1 dz\int_0^z dy(1-z)((1-z)(a-bz) - \tau y(z-y))$$
$$\times\left[\frac{1}{\mathfrak{D}_2(\alpha) + y(1-\alpha)} - \frac{1}{\mathfrak{D}_2(\alpha)}\right] \quad (3.5)$$
$$-\frac{1}{2}\int_0^1 dz\int_0^z dy(a + 5b - 6bz)$$
$$\times[\ln(\mathfrak{D}_2(\alpha) + y(1-\alpha)) - \ln \mathfrak{D}_2(\alpha)],$$

where

$$a = \left(\frac{m_l}{M_W}\right)^2, \quad b = \left(\frac{m_\nu}{M_W}\right)^2,$$

$$\mathfrak{D}_1(\alpha) = \alpha + (a-\alpha)z - bz(1-z) - \tau y(z-y),$$

$$\mathfrak{D}_1 = \mathfrak{D}_1(\alpha = 1) = 1 + (a-1)z - bz(1-z) - \tau y(z-y),$$

$$\mathfrak{D}_2(\alpha) = a + (\alpha - a)z - bz(1-z) - \tau y(z-y),$$

$$\mathfrak{D}_2 = \mathfrak{D}_2(\alpha = 1) = a + (1-a)z - bz(1-z) - \tau y(z-y).$$

Note that the values of the mass parameters of a charge lepton (a) and a neutrino (b) are exactly taken into account in expressions (3.1)–(3.5). The value of the gauge parameter α is arbitrary. The calculations were made for an arbitrary value of q^2.

The contributions from the γ–Z diagrams shown in Fig. 4 to the electric form factor of a neutrino can be obtained on the basis of expansion (2.32) and have the form

$$f_Q^{(j)}(q^2) = \frac{g}{4\cos\theta_W}\frac{A^{(j)}(\alpha, q^2) + B^{(j)}(\alpha, q^2)}{q^2 - M_Z^2}, \quad (3.6)$$
$$j = 7, \ldots, 14.$$

Using the explicit form of functions $A^{(j)}(\alpha, q^2)$ (formulas (2.17)–(2.23)) and $B^{(j)}(\alpha, q^2)$ (formulas (2.24)–(2.30)) and employing relation (3.6), we can derive expressions for the contributions from the γ–Z diagrams for arbitrary values of gauge parameter α and for $q^2 \neq 0$. However, these formulas are cumbersome and are not given here.

4. MAGNETIC FORM FACTOR OF A NEUTRINO

Relation (2.2) representing the general expansion of the neutrino electromagnetic vertex function $\Lambda_\mu(q)$ implies that the neutrino magnetic form factor $f_M(q^2)$ is the coefficient in the term proportional to $i\sigma_{\mu\nu}q^\nu$. In this section, we give the exact expression for $f_M(q^2)$ taking into account the dependence on two mass parameters a and b as well as on parameter α fixing the gauge.

Note that the Feynman diagrams depicted in Fig. 4 make zero contribution to the neutrino magnetic form factor. Thus, the exact expression for the neutrino magnetic form factor has the form

$$f_M(q^2) = \frac{eG_F}{4\pi^2\sqrt{2}} m_\nu \sum_{i=1}^{6} \bar{f}_M^{(i)}(q^2),$$

where coefficients $\bar{f}_M^{(i)}(q^2)$ are the contributions from the corresponding diagrams shown in Fig. 2 to the neutrino magnetic form factor. For this coefficients, we have the following relations:

$$\bar{f}_M^{(1)}(q^2) = \int_0^1 dz \int_0^z dy (2 - 3z + z^2)\frac{1}{D_1}$$

$$-\frac{1}{2}\int_0^1 dz \int_0^z dy (az^2 - bz^2(1-z) - ty(z-y)(2-z)) \times\left[\frac{1}{D_1(\alpha)} - \frac{1}{D_1}\right] \quad (4.1)$$

$$+\frac{1}{2}\int_0^1 dz \int_0^z dy (2-3z)[\ln D_1(\alpha) - \ln D_1],$$

$$\bar{f}_M^{(2)}(q^2) = \frac{1}{2}\int_0^1 dz \int_0^z dy z(a + az - b(1-z))\frac{1}{D_1(\alpha)}, \quad (4.2)$$

$$\bar{f}_M^{(3)}(q^2) = \frac{1}{2}\int_0^1 dz \int_0^z dy (2a - 3az + az^2 - bz(1-z))\frac{1}{D_2(\alpha)}, \quad (4.3)$$

$$\bar{f}_M^{(4)}(q^2) = \frac{1}{2}\int_0^1 dz \int_0^z dy\ z(1+2z)\frac{1}{D_2}$$

$$+\frac{1}{2}\int_0^1 dz \int_0^z dy (b(1-z)^2(z(1-z) - 2y) - ty(z-y)(2y - 3z + z^2) - 2ty) \times\left[\frac{1}{D_2(\alpha) + y(1-\alpha)} - \frac{1}{D_2}\right]$$

$$+\frac{1}{2}\int_0^1 dz \int_0^z dy (-2 + 9z - 4z^2 - 6y) \times [\ln(D_2(\alpha) + y(1-\alpha)) - \ln D_2] \quad (4.4)$$

$$-\frac{t}{4}\int_0^1 dz \int_0^z dy (bz(1 - 3z + z^2 + z^3) - ty(z-y)(2 - z - z^2)) \times\left[\frac{1}{D_2} + \frac{1}{D_2(\alpha)} - \frac{2}{D_2(\alpha) + y(1-\alpha)}\right]$$

$$+\frac{t}{8}\int_0^1 dz \int_0^z dy (8 - 13z + 3z^2) \times [\ln D_2 + \ln D_2(\alpha) - 2\ln(D_2(\alpha) + y(1-\alpha))],$$

$$\bar{f}_M^{(5)+(6)}(q^2) = \int_0^1 dz \int_0^z dy y \frac{1}{D_2(\alpha) + y(1-\alpha)}$$

$$+\frac{1}{2}\int_0^1 dz \int_0^z dy ((a - bz)(1-z)^2 + ty(z-y)(1-z)) \times\left[\frac{1}{D_2(\alpha) + y(1-\alpha)} - \frac{1}{D_2(\alpha)}\right] \quad (4.5)$$

$$+\frac{1}{2}\int_0^1 dz \int_0^z dy (2-3z)[\ln(D_2(\alpha) + y(1-\alpha)) - \ln D_2(\alpha)],$$

here,

$$D_1(\alpha) = \alpha + (a-\alpha)z - bz(1-z) + ty(z-y),$$

$$D_1 = D_1(\alpha = 1) = 1 + (a-1)z - bz(1-z) + ty(z-y),$$

$$D_2(\alpha) = a + (\alpha - a)z - bz(1-z) + ty(z-y),$$

$$D_2 = D_2(\alpha = 1) = a + (1-a)z - bz(1-z) + ty(z-y),$$

$t = -q^2/M_W^2$.

Let us analyze the asymptotic behavior of the integrals contained in the contributions from proper vertices diagrams to $f_M(q^2)$ for large positive values of t. For example, let us consider the following integral for $t \longrightarrow \infty$:

$$J(t) = t\int_0^z dy \frac{y}{D_2(\alpha)} = \int_0^z dy \frac{y}{(y-y_2)(y_1-y)}; \quad (4.6)$$

here,

$$y_1 = z + \frac{D}{zt} + \ldots, \quad y_2 = -\frac{D_\alpha}{zt} + \ldots, \quad (4.7)$$

and

$$\begin{aligned} D_\alpha &= a + (\alpha - a)z - bz(1-z), \\ D &= D_{\alpha=1} = a + (1-a)z - bz(1-z). \end{aligned} \quad (4.8)$$

Carrying out elementary integration, we obtain

$$J(t) \longrightarrow \ln t - \ln D. \quad (4.9)$$

In formulas (4.6)–(4.8), we omitted terms proportional to $1/t$ and $(\ln t)/t$, which are infinitely small for large positive values of t. The remaining integrals can be estimated similarly. Finally, we find that

$$\bar{f}_M(t) = \sum_{i=1}^{6} \bar{f}_M^{(i)}(t) \longrightarrow 0 \text{ as } t \longrightarrow \infty.$$

The asymptotic behavior of the magnetic form factor for large negative values of q^2 described here is in accordance with the general Weinberg theorem [24]. However, the case of the magnetic form factor of a massive neutrino has never been analyzed before. It should be noted that, in deriving relations (4.6)–(4.8), we assumed that $\alpha < \infty$. Thus, the obtained result ($\bar{f}_M(t) \longrightarrow 0$ as $t \longrightarrow \infty$) is valid for any gauge except the unitary one. The value of $\bar{f}_M(t \longrightarrow \infty)$ may differ from zero if we first set $\alpha = \infty$ and then proceed to the limit $t \longrightarrow +\infty$. The behavior of magnetic form factors in the framework of the Weinberg–Salam model in the unitary gauge is analyzed, for example, in [14]. Using explicit expressions for the magnetic form factor of a massive neutrino for an arbitrary value of gauge parameter α, Fig. 5 shows the behavior of function $\bar{f}_M(t)$ in various gauges for a wide range of t values: $0 \le t \le 5 \times 10^{-4}$. It can be seen that the magnetic form factor becomes independent of the choice of the gauge at $t = 0$, which corresponds to a photon on the mass surface. The value of $f_M(t = 0)$ is equal to the neutrino magnetic moment. The fact that the magnetic moment of the massive neutrino is independent of the choice of the gauge was proved by direct calculation in our previous publication [13].

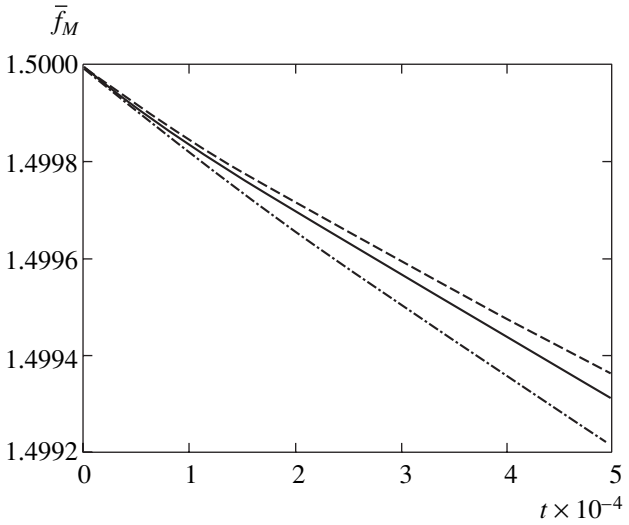


Fig. 5. Magnetic form factor of a massive neutrino as a function of t for various values of the gauge parameter. The dashed curve corresponds to $\alpha = 100$, the solid curve to the 't Hoft–Feynman gauge ($\alpha = 1$), and the dot-and-dash curve, to $\alpha = 0.1$.

5. ANAPOLE FORM FACTOR OF A NEUTRINO

In this section, we consider the anapole form factor of a massive neutrino. Using the results obtained in Section 2 for various contributions to the neutrino vertex function $\Lambda_\mu(q)$, we single out in formulas (2.3)–(2.15) the coefficients proportional to $(q^2\gamma_\mu - \not{q}q_\mu)\gamma_5$, which are, in accordance with the expansion in relation (2.2), the corresponding contributions to the anapole form factor $f_A(q^2)$. It should be noted that, in separating similar terms in the contributions from each of the diagrams depicted in Figs. 2 and 4, we inevitably obtain additional terms proportional to matrix $\gamma_\mu\gamma_5$. Consequently, it is necessary to make sure that the corresponding form factor has zero value even at $q^2 \neq 0$. This problem was also touched upon in Section 2.

We will first consider the contributions from single-loop proper vertices diagrams (see Fig. 2) to the neutrino anapole form factor. Integrating with respect to momenta of virtual particles (see monograph [23]), we obtain exact expression for the contributions of these diagrams to the anapole form factor of a massive neutrino in terms of definite integrals,

$$f_A(q^2) = \frac{eG_F}{4\pi^2\sqrt{2}} \sum_{i=1}^{6} \bar{f}_A^{(i)}(q^2),$$

where

$$\bar{f}_A^{(1)}(q^2) = -\frac{1}{2}\int_0^1 dz\int_0^z dy((2-z-z^2)+4y(z-y))\frac{1}{\mathfrak{D}_1}$$

$$-\frac{1}{4}\int_0^1 dz\int_0^z dy[(a+b(1-z))z^2$$

$$-4y(z-y)(a-b)-2\tau y(z-y)(2-z)]\left[\frac{1}{\mathfrak{D}_1(\alpha)}-\frac{1}{\mathfrak{D}_1}\right] \quad (5.1)$$

$$-\frac{1}{4}\int_0^1 dz\int_0^z dy(2-3z)[\ln\mathfrak{D}_1(\alpha)-\ln\mathfrak{D}_1],$$

$$\bar{f}_A^{(2)}(q^2) = -\frac{a-b}{4}$$

$$\times\int_0^1 dz\int_0^z dy(z(1-z)+4y(z-y))\frac{1}{\mathfrak{D}_1(\alpha)}, \quad (5.2)$$

$$\bar{f}_A^{(3)}(q^2) = \frac{a-b}{4}\int_0^1 dz\int_0^z dy(z-2y)^2\frac{1}{\mathfrak{D}_2(\alpha)}, \quad (5.3)$$

$$\bar{f}_A^{(4)}(q^2) = -\frac{1}{4}\int_0^1 dz\int_0^z dy(z(3-2z)+8y(z-y))\frac{1}{\mathfrak{D}_2}$$

$$-\frac{1}{4}\int_0^1 dz\int_0^z dy[bz(1-z^2)(3-z)$$

$$-4b(1-z)(z-y(1+z))(z-y)$$

$$-\tau(4y^3-4y^2z-2y+yz^2-4y^2+yz)(z-y)]$$

$$\times\left[\frac{1}{\mathfrak{D}_2(\alpha)+y(1-\alpha)}-\frac{1}{\mathfrak{D}_2}\right]$$

$$+\frac{1}{4}\int_0^1 dz\int_0^z dy(6-3z-4z^2+12y+16y(z-y)) \quad (5.4)$$

$$\times[\ln(\mathfrak{D}_2(\alpha)+y(1-\alpha))-\ln\mathfrak{D}_2]$$

$$-\frac{1}{4}\int_0^1 dz\int_0^z dy[b(1-z^2)+\tau(z+2y(z-y))]$$

$$\times[\ln\mathfrak{D}_2+\ln\mathfrak{D}_2(\alpha)-2\ln(\mathfrak{D}_2(\alpha)+y(1-\alpha))]$$

$$+\frac{3}{4}\int_0^1 dz\int_0^z dy[\mathfrak{D}_2\ln\mathfrak{D}_2+\mathfrak{D}_2(\alpha)\ln\mathfrak{D}_2(\alpha)$$

$$-2(\mathfrak{D}_2(\alpha)+y(1-\alpha))\ln(\mathfrak{D}_2(\alpha)+y(1-\alpha))],$$

$$\bar{f}_A^{(5)+(6)}(q^2) = -\frac{1}{2}\int_0^1 dz\int_0^z dyy\frac{1}{\mathfrak{D}_2(\alpha)+y(1-\alpha)}$$

$$+\frac{1}{4}\int_0^1 dz\int_0^z dy[a((1+z)^2-4y(1+z-y))$$

$$-b(z(1+z)^2-2y((1+z)^2-2y))$$

$$+\tau y(z-y)(1+z-2y)]\left[\frac{1}{\mathfrak{D}_2(\alpha)+y(1-\alpha)}-\frac{1}{\mathfrak{D}_2(\alpha)}\right] \quad (5.5)$$

$$-\frac{1}{4}\int_0^1 dz\int_0^z dy(2+3z-6y)$$

$$\times[\ln(\mathfrak{D}_2(\alpha)+y(1-\alpha))-\ln\mathfrak{D}_2(\alpha)].$$

Note that the values of the mass parameters of a charged lepton (a) and a neutrino (b) are taken into account exactly in expressions (5.1)–(5.5). The value of gauge parameter α is arbitrary. All calculations were made for an arbitrary value of q^2.

The contributions from the γ–Z diagrams shown in Fig. 4 to the neutrino anapole form factor can be derived using expansion (2.32) and have the form

$$f_A^{(j)}(q^2) = \frac{g}{4\cos\theta_W}\frac{1}{q^2-M_Z^2}$$

$$\times\left\{\frac{A^{(j)}(\alpha,q^2)}{q^2}+(1-\alpha_Z)\frac{B^{(j)}(\alpha,q^2)}{q^2-\alpha_Z M_Z^2}\right\}, \quad (5.6)$$

$$j = 7,\ldots,14.$$

Using the explicit form of functions $A^{(j)}(\alpha, q^2)$ (formulas (2.17)–(2.23)) and $B^{(j)}(\alpha, q^2)$ (formulas (2.24)–(2.30)) and employing relation (5.6), we can derive expressions for the contributions from the γ–Z diagrams for arbitrary values of gauge parameters α and α_Z and for $q^2 \neq 0$. However, these formulas are cumbersome and will not be given here.

Anapole Moment

Let us consider the anapole moment of a massive neutrino. We have obtained the contributions from the proper vertices diagrams (formulas (5.1)–(5.5)) as well as from the γ–Z diagrams (5.6) to the neutrino anapole

form factor for an arbitrary value of q^2. Since the anapole moment is a static electromagnetic parameter of the neutrino, the value of q^2 should be set equal to zero in the formulas under consideration.

In the case of a zero-mass neutrino, the value of the anapole moment is connected with the charge radius via the relation (see, for example, [21])

$$a_\nu = \frac{1}{6}\langle r_\nu^2\rangle.$$

However, in the case of a massive particle, this simple relationship is violated for the reasons described in Section 2.

In the expression for the anapole moments, the contributions come both from the proper vertices diagrams depicted in Fig. 2 and from the γ–Z diagrams shown in Fig. 4. Thus, the complete expression for the anapole moment has the form

$$a_\nu = \frac{eG_F}{4\pi^2\sqrt{2}}\left\{\sum_{i=1}^{6}\bar{a}^{(i)}(a, b, \alpha) + \sum_{j=7}^{14}\bar{a}^{(j)}(a, \alpha)\right\}.$$

The contributions $\bar{a}^{(i)}(a, b, \alpha)$ from the proper vertices diagrams can be written in the form

$$\bar{a}^{(1)}(a, b, \alpha) = -\frac{1}{6}\int_0^1 dz(2 + 3z - 6z^2 + z^3)\frac{1}{D}$$
$$-\frac{1}{12}\int_0^1 dz(1-z)^3(a + 2b + 3bz)\left[\frac{1}{D_\alpha} - \frac{1}{D}\right] \qquad (5.7)$$
$$+\frac{1}{4}\int_0^1 dz(1-z)(1 - 4z + 3z^2)[\ln D_\alpha - \ln D],$$

$$\bar{a}^{(2)}(a, b, \alpha) = -\frac{a-b}{12}\int_0^1 dz(2 - 3z + z^3)\frac{1}{D_\alpha}, \qquad (5.8)$$

$$\bar{a}^{(3)}(a, b, \alpha) = \frac{a-b}{12}\int_0^1 dz z^3\frac{1}{D_\alpha}, \qquad (5.9)$$

$$\bar{a}^{(4)}(a, b, \alpha) = -\frac{1}{12}\int_0^1 dz z^2(9 - 2z)\frac{1}{D}$$
$$-\frac{b}{4}\int_0^1 dz\int_0^z dy[z(1-z^2)(3-z)$$
$$-4(1-z)(z - y(1+z))(z-y)]\left[\frac{1}{D_\alpha + y(1-\alpha)} - \frac{1}{D}\right]$$
$$+\frac{1}{4}\int_0^1 dz\int_0^z dy(6 - 3z - 4z^2 + 12y + 16y(z-y)) \qquad (5.10)$$
$$\times[\ln(D_\alpha + y(1-\alpha)) - \ln D]$$

$$-\frac{b}{4}\int_0^1 dz\int_0^z dy(1-z^2)[\ln D + \ln D_\alpha - 2\ln(D_\alpha + y(1-\alpha))]$$
$$+\frac{3}{4}\int_0^1 dz\int_0^z dy[D\ln D + D_\alpha\ln D_\alpha$$
$$-2(D_\alpha + y(1-\alpha))\ln(D_\alpha + y(1-\alpha))],$$

$$\bar{a}^{(5)+(6)}(a, b, \alpha) = -\frac{1}{2}\int_0^1 dz\int_0^z dy y\frac{1}{D_\alpha + y(1-\alpha)}$$
$$+\frac{1}{4}\int_0^1 dz\int_0^z dy[a((1+z)^2 - 4y(1+z-y))$$
$$-b(x(1+z)^2 - 2y((1+z)^2 - 2y))] \qquad (5.11)$$
$$\times\left[\frac{1}{D_\alpha + y(1-\alpha)} - \frac{1}{D_\alpha}\right] - \frac{1}{4}\int_0^1 dz\int_0^z dy(2 + 3z - 6y)$$
$$\times[\ln(D_\alpha + y(1-\alpha)) - \ln D_\alpha].$$

It should be noted that, as in the case of a zero-mass neutrino, the contributions from proper vertices diagrams to the anapole moment are finite.

To obtain the contributions from the γ–Z diagrams to the anapole moment, it is convenient to use formula (2.32), setting $q^2 = 0$ in this case. Thus, the expressions for

$$a^{(j)}(a, \alpha) = \frac{eG_F}{4\pi^2\sqrt{2}}\bar{a}^{(j)}(a, \alpha)$$

assume the form

$$a^{(j)}(a, \alpha) = \frac{1}{M_Z^2}Q^{(j)} - \frac{g}{4M_Z^2\cos\theta_W}\frac{A^{(j)}(q^2)}{q^2}\bigg|_{q^2\to 0}, \qquad (5.12)$$
$$j = 7, \ldots, 14,$$

where the contributions to the neutrino electric charge $Q^{(j)}$ can be derived from expressions (2.24)–(2.30) and (3.6) (see also our previous publication [13]). While deriving relation (5.12), we assumed that $\alpha_Z = \infty$. Using

formula (5.12), we obtain the following expressions for $\bar{a}^{(j)}(a, \alpha)$:

$$\bar{a}^{(7)}(a,\alpha) = \cos^2\theta_W\left\{\cos^2\theta_W\left[\omega\left(3+\frac{3}{4}\alpha(1+\alpha)\right)\right.\right.$$
$$\left.-1-\frac{5\alpha}{8}-\frac{5\alpha^2}{8}-\frac{3\alpha^3}{4}\frac{\ln\alpha}{1-\alpha}\right]$$
$$-\frac{1}{2}\left[\omega\left(-\frac{14}{3}+\alpha\right)-\frac{1}{18(1-\alpha)^3}\right. \quad (5.13)$$
$$\times(11-54\alpha+54\alpha^2-2\alpha^3-9\alpha^4$$
$$\left.\left.-18\alpha^2\ln\alpha-12\alpha^3\ln\alpha+18\alpha^4\ln\alpha)\right]\right\},$$

$$\bar{a}^{(8)}(a,\alpha) = \sin^2\theta_W\left\{\cos^2\theta_W\left[\omega\frac{3+\alpha}{4}-\frac{5+\alpha}{8}\right.\right.$$
$$\left.-\frac{\alpha}{2}\left(1+\frac{\alpha}{2}\right)\frac{\ln\alpha}{1-\alpha}\right]-\frac{1}{18(1-\alpha)^3} \quad (5.14)$$
$$\left.\times(11-18\alpha+9\alpha^2-2\alpha^3+6\ln\alpha)\right\},$$

$$\bar{a}^{(9)}(a,\alpha) = (\cos^2\theta_W-\sin^2\theta_W)\cos^2\theta_W$$
$$\times\frac{1}{2}\{-\omega\alpha+\alpha-\alpha\ln\alpha\}, \quad (5.15)$$

$$\bar{a}^{(10)}(a,\alpha) = \cos^4\theta_W$$
$$\times\left\{-\frac{3}{4}\omega(3+\alpha^2)+\frac{3}{8}+\frac{5\alpha^2}{8}-\frac{3}{4}\alpha^2\ln\alpha\right\}, \quad (5.16)$$

$$\bar{a}^{(11)+(12)}(a,\alpha) = \cos^2\theta_W$$
$$\times\frac{1}{2}\left\{\cos^2\theta_W(-\omega\alpha+\alpha-\alpha\ln\alpha)-\frac{1}{3}(\omega+\ln\alpha)\right\}, \quad (5.17)$$

$$\bar{a}^{(13)}(a,\alpha) = (\sin^2\theta_W-\cos^2\theta_W)$$
$$\times\frac{1}{2}\left\{\cos^2\theta_W(-\omega\alpha+\alpha-\alpha\ln\alpha)+\frac{1}{6}(\omega+\ln\alpha)\right\}, \quad (5.18)$$

$$\bar{a}^{(14)}(a,\alpha) = -\left\{\omega\left(-1-\frac{28}{9}\sin^2\theta_W\right)\right.$$
$$\left.+\frac{1}{3}\sum_f Q_f\left(\pm\frac{1}{2}-2Q_f\sin^2\theta_W\right)\ln\left(\frac{m_f}{M_W}\right)^2\right\}. \quad (5.19)$$

Expressions (5.13)–(5.19) are diverging; consequently, the final form of these formulas depends on the method for regularizing Feynman integrals (2.9)–(2.15). This circumstance can be used to explain a certain difference between relations (5.13)–(5.19) and the corresponding contributions to the charge radius, which were derived in [16].

The direct calculation performed here gives diverging expressions for the anapole moment of a massive neutrino. It should be recalled in this connection that the corresponding corrections to the electromagnetic vertex function can be treated from the standpoint of radiative corrections to the expression for a physical process (e.g., neutrino scattering by a charged lepton). It is quite obvious that the scattering cross section, which is a measurable quantity, must be finite and independent of the choice of the gauge. It is precisely this approach, which was developed in [17, 21, 25, 26] for the case of a zero-mass neutrino, that formed the basis of the definition of the electroweak anapole moment and the electroweak charge radius. Note that, in addition to the Feynman diagrams considered here, the corrections to the electromagnetic vertex function of a charged lepton should also be taken into account in studying the radiative corrections to scattering. Moreover, the so-called "box" diagram, in which the neutrino and the lepton exchange two virtual bosons in the course of the interaction, will also affect the scattering process. Detailed calculations of the corresponding diagrams for the case of a zero-mass neutrino are given in the recent publications [25, 26]. It would be interesting to demonstrate that, for a massive neutrino also, the diverging terms depending on the gauge parameters will cancel out. However, in the case of a massive neutrino, this problem is not trivial and requires an additional independent analysis.

6. ELECTROMAGNETIC CHARACTERISTICS OF A NEUTRINO IN THE CASE OF MIXING BETWEEN DIFFERENT GENERATIONS

It was noted in Section 1 that contemporary experimental data speak in favor of mixing existing between different flavors of neutrinos. In the study of the electromagnetic properties of neutrinos, this is manifested in the existence of a transitional electromagnetic dipole moment (and also transitional form factors) of a neu-

trino. Indeed, formula (2.1) in the present case assumes the form

$$\langle \nu(p')\left|J_\mu^{\mathrm{EM}}\right|\nu(p)\rangle = \sum_{\alpha\beta} \bar{u}_\beta(p')\Lambda_\mu^{\beta\alpha}(q)u_\alpha(p),$$

where summation is carried out over neutrino types (ν_e, ν_μ and ν_τ). By analogy with expansion (2.2), we conclude that the most general expression for vertex function $\Lambda_\mu^{\beta\alpha}(q)$ is

$$\begin{aligned}\Lambda_\mu^{\beta\alpha}(q) = {} & f_Q^{\beta\alpha}\gamma_\mu + f_M^{\beta\alpha} i\sigma_{\mu\nu}q^\nu - f_E^{\beta\alpha}\sigma_{\mu\nu}q^\nu\gamma_5 \\ & + f_A^{\beta\alpha}(q^2\gamma_\mu - q_\mu \not{q})\gamma_5,\end{aligned} \tag{6.1}$$

where the quantities $f_I^{\beta\alpha}$ ($I = Q, M, E, A$) have the meaning of transitional form factors.

By way of an example, we analyze the minimally extended standard model with mixing between different flavors of charged leptons and neutrinos. The interaction Lagrangian and the Feynman rules are given in [22]. First, we consider the situation when $m_{\nu_e} = m_{\nu_\mu} = m_{\nu_\tau} = m_\nu$. Such a choice of parameters corresponds to three flavors of neutrinos degenerate in mass. In this case, we can prove that expressions for $f_I^{\beta\alpha}$ have the form

$$f_I^{\beta\alpha} = \sum_{l = e, \mu, \tau} U_{\beta l}\overset{*}{U}_{\alpha l} f_I^{(l)}, \tag{6.2}$$

where $U_{\beta l}$ is the unitary matrix describing mixing between different generations of leptons and neutrinos (see review [22]) and $f_I^{(l)} = f_I(a_l, b, q^2, \ldots)$ are the expressions for form factors derived in Sections 3–5 of this paper; $a_l = (m_l/M_W)^2$.

In actual practice, the case of nearly degenerate neutrino masses is not ruled out by the available experimental data (see [27, 28]). Considering the case of mass-degenerate neutrinos, we in fact introduce two additional (apart from parameters $a = (m_l/M_W)^2$ and $b = (m_\nu/M_W)^2$ used above) mass parameters $c_i = (\Delta m_i/M_W)^2$, where Δm_i are two independent values of the mass differences between three flavored neutrinos, and carry out the expansion in these parameters in subsequent calculations, assuming that $c \ll a, b$.

If we set $q^2 = 0$ in formula (6.2), this will lead to expressions for transient charges, which determine the static electromagnetic properties of three types of neutrino. It is interesting to note that the structure of relation (6.2) and the results obtained in out previous publications [13] imply that the values of neutrino transient charges $Q_{\beta\alpha} = f_Q^{\beta\alpha}(q^2 = 0)$ are identically equal to zero for an arbitrary value of gauge parameter α.

In the case of nondegenerate neutrino masses, apart from the three form factors considered in this section, relation (6.1) also acquires a transitional electric dipole form factor that is identically equal to zero for neutrinos with degenerate masses. For this reason, we will give the expression for this form factor only. The simplest expression for the electric dipole form factor is obtained in the 't Hoft–Feynman gauge. Note that the contributions to this form factor comes only from proper vertices diagrams. On the basis of expressions (2.3)–(2.7) (substituting m_β for m_ν in the extreme left parentheses and m_α for m_ν in the extreme right parentheses in formulas (2.4) and (2.5), where $m_{\beta,\alpha}$ are the flavor masses of the final and the initial neutrino states), averaging over the initial and final neutrino states, we find that

$$f_E^{\beta\alpha}(q^2) = \frac{eG_\mathrm{F}}{4\pi^2\sqrt{2}}\sum_{i=1}^{6}\left(\sum_{l = e, \mu, \tau} U_{\beta l}\overset{*}{U}_{\alpha l}\bar{E}_{\beta\alpha}^{(i)}(q^2, a_l)\right).$$

Here, the contributions from each Feynman diagram have the form

$$\begin{aligned}\bar{E}_{\beta\alpha}^{(1)} = {} & -\frac{i}{2}\int_0^1 dx \int_0^{1-x} dy \frac{1}{D_{\beta\alpha}^{(1)}} \\ & \times [(m_\beta - m_\alpha)(2 - 3(x+y) + (x+y)^2) \\ & + (m_\beta + m_\alpha)(x - y - (x^2 - y^2))],\end{aligned} \tag{6.3}$$

$$\begin{aligned}\bar{E}_{\beta\alpha}^{(2)} = {} & -\frac{i}{4}\int_0^1 dx \int_0^{1-x} dy \frac{1}{D_{\beta\alpha}^{(1)}} \\ & \times \{a_l[(m_\beta - m_\alpha)(x + y + (x+y)^2) \\ & + (m_\beta + m_\alpha)(x - y - (x^2 - y^2))] \\ & + \sqrt{b_\alpha b_\beta}[(m_\beta - m_\alpha)((x+y) - (x+y)^2) \\ & - (m_\beta + m_\alpha)((x-y) - (x^2 - y^2))]\},\end{aligned} \tag{6.4}$$

$$\begin{aligned}\bar{E}_{\beta\alpha}^{(3)} = {} & \frac{i}{4}\int_0^1 dx \int_0^{1-x} dy \frac{1}{D_{\beta\alpha}^{(2)}}(x + y - 1) \\ & \times [(m_\beta - m_\alpha)(2a_l + (x+y)(\sqrt{b_\alpha b_\beta} - a_l)) \\ & - (m_\beta + m_\alpha)(x - y)(\sqrt{b_\alpha b_\beta} - a_l)],\end{aligned} \tag{6.5}$$

$$\begin{aligned}\bar{E}_{\beta\alpha}^{(4)} = {} & -\frac{i}{4}\int_0^1 dx \int_0^{1-x} dy \frac{1}{D_{\beta\alpha}^{(2)}} \\ & \times [(m_\beta - m_\alpha)(x + y + 2(x+y)^2) \\ & + (m_\beta + m_\alpha)(x - y - 2(x^2 - y^2))],\end{aligned} \tag{6.6}$$

$$\bar{E}^{(5)+(6)}_{\beta\alpha} = \frac{i}{4}\int_0^1 dx \int_0^{1-x} dy \frac{1}{D^{(2)}_{\beta\alpha}} \times [(m_\beta - m_\alpha)(x+y) + (m_\beta + m_\alpha)(x-y)], \quad (6.7)$$

and

$$D^{(1)}_{\beta\alpha} = 1 + (a_l - 1)(x+y) - \frac{b_\beta + b_\alpha}{2}(x+y)(1-x-y) - \tau xy + \frac{b_\beta - b_\alpha}{2}(x - y - (x^2 - y^2)),$$

$$D^{(2)}_{\beta\alpha} = a_l + (1 - a_l)(x+y) - \frac{b_\beta + b_\alpha}{2}(x+y)(1-x-y) - \tau xy + \frac{b_\beta - b_\alpha}{2}(x - y - (x^2 - y^2)),$$

$b_\alpha = (m_\alpha/M_W)^2$. We have derived formulas (6.3)–(6.7) using the fact that the relation

$$\bar{u}_\beta(p')(p'_\mu + p_\mu)\gamma_5 u_\alpha(p) = \bar{u}_\beta(p')[(m_\beta - m_\alpha)\gamma_\mu\gamma_5 - i\sigma_{\mu\nu}\gamma_5 q^\nu]u_\alpha(p)$$

also holds for a neutrino in the mass shell.

It follows from formulas (6.3)–(6.7) that, in the case of degenerate neutrino masses (i.e., for $m_\beta = m_\alpha$), the contributions from each Feynman diagram to the electric dipole form factor are identically equal to zero.

On the basis of relations (6.3)–(6.7), we can derive the expressions for transitional electric dipole moments of neutrinos. To this end, we must set $\tau = 0$ in the formulas under investigation. The simplest expressions for the electric dipole moments are obtained for light neutrinos ($b_\alpha \ll 1$). Expanding the integrands in formulas (6.3)–(6.7) in parameter b_α and integrating with respect to Feynman parameters x and y, we obtain

$$\sum_{i=1}^{6} \bar{E}^{(i)}_{\beta\alpha}(q^2 = 0, a_l) = -\frac{i}{24(1-a_l)^4}(m_\beta - m_\alpha) \times (12 - 52a_l + 81a_l^2 - 48a_l^3 + 7a_l^4 - 6a_l^2 \ln a_l + 12a_l^3 \ln a_l). \quad (6.8)$$

Proceeding from the fact that charged leptons must be much lighter than a W boson (i.e., $a_l \ll 1$), we obtain from relation (6.8) the final expression for transitional electric dipole moments of the neutrino in the form

$$d_{\beta\alpha} = f_E^{\beta\alpha}(q^2 = 0) = \frac{eG_F}{8\pi^2\sqrt{2}} i(m_\alpha - m_\beta). \quad (6.9)$$

Note that this relation does not contain the dependence on the mixing angle since we disregard the masses of charged leptons. It is obvious from formula (6.9) that, in the minimally extended standard model with mixing between different flavors of charged leptons and neutrinos, the neutrino electric dipole moments that are diagonal in flavors are equal to zero, while nondiagonal elements are proportional to the difference in the neutrino flavor masses.

7. CONCLUSIONS

We have investigated the electromagnetic vertex form factors of a massive Dirac neutrino in the framework of the minimally extended standard model supplemented with a SU(2)-singlet right-handed neutrino. In all calculations, we have exactly taken into account the masses of a charged lepton and a neutrino. Calculations were made in an arbitrary R_ξ gauge, which makes it possible to analyze the dependence of the results on the gauge parameters of both W and Z bosons. It was found in Section 2 that, for a certain choice of gauge parameters, all electromagnetic form factors of the neutrino become finite (i.e., contain no ultraviolet divergence). This statement has been proved in the one-loop approximation. An analogous property of the electromagnetic vertex function can be formulated in the framework of quantum electrodynamics. For a definite choice of the photon gauge parameter ($\alpha_\gamma = 3$), the electron electromagnetic vertex function in the one-loop approximation does not contain infrared divergence. In Sections 3 and 4, the contributions from all Feynman diagrams to the charge and magnetic form factors, which exactly take into account the dependence on mass parameters a and b as well as on the gauge parameter α, have been determined for the first time. The asymptotic behavior of the magnetic form factor of a massive neutrino is investigated for $q^2 \longrightarrow -\infty$ and it is found that $f_M(q^2) \longrightarrow 0$ in this case. The anapole form factor and the anapole moment of a massive neutrino are considered in Section 5 for an arbitrary value of gauge parameter α. It is also found that, like in the case of a zero-mass particle, the anapole moment of a massive neutrino is a diverging quantity and depends of the choice of the gauge. The transient electromagnetic form factors of the neutrino are studied in Section 6 in the framework of the minimally extended standard model permitting mixing between different flavors of charged leptons and neutrinos. Using the results obtained in Sections 3–5, we have obtained for the first time the explicit expressions for the transient charge, magnetic, and anapole form factors for neutrino states degenerate in flavor masses. It is shown that transient electric charges are identically equal to zero. For the case of neutrino states nondegenerate in masses, an exact expression for the transitional electric dipole form factor is obtained in the 't Hoft–Feynman gauge. This form factor is identically equal to zero for mass-degenerate neutrinos. Moreover, the expression for the transitional electric dipole moment was also found in this gauge. It was found that the transitional electric

dipole moment is proportional to the difference in the masses of the initial and final neutrino states.

REFERENCES

1. Q. R. Ahmad, R. C. Allen, T. C. Andersen, *et al.*, Phys. Rev. Lett. **89**, 011301 (2002).
2. Y. Fukuda, T. Hayakawa, E. Ichihara, *et al.*, Phys. Rev. Lett. **82**, 2644 (1999).
3. M. H. Ahn, S. Aoki, H. Bhang, *et al.*, Phys. Rev. Lett. **90**, 041801 (2003).
4. B. Kayser, Phys. Rev. D **26**, 1662 (1982).
5. W. Bardeen, R. Gastmans, and B. Lautrup, Nucl. Phys. B **46**, 319 (1972).
6. B. W. Lee and R. E. Shrock, Phys. Rev. D **16**, 1444 (1977).
7. K. Fujikawa and R. E. Shrock, Phys. Rev. Lett. **45**, 963 (1980).
8. W. J. Marciano and A. Sirlin, Phys. Rev. D **22**, 2695 (1980).
9. S. Sakakibara, Phys. Rev. D **24**, 1149 (1981).
10. R. E. Shrock, Nucl. Phys. B **206**, 359 (1982).
11. J. L. Lucio Martínez, A. Rosado, and A. Zepeda, Phys. Rev. D **29**, 1539 (1984).
12. L. G. Cabral-Rosetti, J. Bernabéu, J. Vidal, *et al.*, Eur. Phys. J. C **12**, 633 (2000).
13. M. Dvornikov and A. Studenikin, Phys. Rev. D **69**, 073001 (2004).
14. K. Fujikawa, B. W. Lee, and A. I. Sanda, Phys. Rev. D **6**, 2923 (1972).
15. J. E. Kim, Phys. Rev. D **14**, 3000 (1976).
16. J. L. Lucio, A. Rosado, and A. Zepeda, Phys. Rev. D **31**, 1091 (1985).
17. J. Bernabéu, L. G. Cabral-Rosetti, J. Papavassiliou, *et al.*, Phys. Rev. D **62**, 113012 (2000).
18. H. Czyz, K. Kolodziej, M. Zralek, *et al.*, Can. J. Phys. **66**, 132 (1988).
19. V. Dubovik and V. Kuznetsov, Int. J. Mod. Phys. A **13**, 5257 (1998).
20. E. N. Bukina, V. M. Dubovik, and V. E. Kuznetsov, Yad. Fiz. **61**, 1129 (1998) [Phys. At. Nucl. **61**, 1035 (1998)].
21. A. Rosado, Phys. Rev. D **61**, 013001 (2000).
22. K. Aoki, Z. Hioki, R. Kawabe, *et al.*, Prog. Theor. Phys. Suppl. **73**, 1 (1982).
23. N. N. Bogolyubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields*, 4th ed. (Nauka, Moscow, 1984; Wiley, New York, 1980).
24. C. Itzykson and J. B. Zuber, *Introduction to Quantum Field Theory* (McGraw-Hill, New York, 1980; Mir, Moscow, 1984), Vol. 2.
25. J. Bernabéu, J. Papavassiliou, and J. Vidal, Phys. Rev. Lett. **89**, 101802(E) (2002); **89**, 229902 (2002).
26. J. Bernabéu, J. Papavassiliou, and J. Vidal, Nucl. Phys. B **680**, 450 (2004).
27. J. I. Illana and M. Masip, hep-ph/0307393.
28. S. M. Bilenky, Proc. R. Soc. London, Ser. A **460**, 403 (2004).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 99, No. 2, 2004, pp. 270–278.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 2, 2004, pp. 305–314.

ATOMS, SPECTRA, RADIATION

The Shape of Saturated Absorption Resonance in Atoms with Degenerate Levels and Light Pressure Effect

A. A. Chernenko* and A. V. Shishaev

Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia

**e-mail: chernen@isp.nsc.ru*

Received November 18, 2003

Abstract—Anomalous behavior of the shape of saturated absorption resonance on the transition $1s_5 \longrightarrow 2p_8$ ($J = 2 \longrightarrow J = 2$) in the Ne atom is experimentally observed. The shape of the saturation resonance on transitions with degenerate excited states of atoms is analyzed numerically and the reasons for the anomalous behavior of the resonance shape and the formation of its doublet spectral structure are established. The effect of the resonance light pressure on the amplitude and frequency properties of resonance is investigated. It is shown that the asymmetry of the doublet splitting of the resonance is associated with the effect of resonance light pressure.

1. INTRODUCTION

Degenerate atomic systems with a long-lived lower (ground, or metastable) state remain interesting as an object for studying processes of their interaction with high-intensity laser radiation not only due to rich spectroscopic manifestations emerging in this case, but also possible practical applications for solving a number of fundamental and applied problems.

The saturated absorption spectra of such Λ systems under the conditions of optical orientation in strong laser fields were studied experimentally on atomic transitions in Na and Ne [1–3]. An important result of these experiments was the observation of a dip in saturated absorption resonances at high strengths of the orienting field. A peculiarity in the dips was their anomalously small width for large values of the saturation parameter and asymmetric arrangement relative to the center of the resonance line.

The spectral singularities observed in [3] stimulated theoretical analysis of spectroscopic manifestations emerging during the interaction of degenerate metastable atomic systems with an orienting field of arbitrary intensity [4, 5]. Analysis carried out for transitions between levels with identical values of the total angular momenta $J = 2$ revealed that the shape of the saturated absorption resonance in this case is determined by a number of factors that are far from obvious and lead to the formation of dips and peaks of various amplitude and width as well as to their superposition. An important result was the discovery of a strong dependence of these parameters on the value of the branching factor characterizing the relative probabilities of relaxation channels of the excited state.

An important property of degenerate atomic systems as applied to resonant interaction with laser radiation is the formation of several two-level M subsystems, which are coupled via spontaneous and induced transitions (M is the magnetic quantum number of the level). As a result, the system of kinetic equations describing the process of interaction is one of $2M$th order even if we disregard correlations between magnetic sublevels. For $M \geq 2$, it is very difficult to obtain an exact analytic solution to such systems of equations. For this reason, such systems are solved using approximations in a number of parameters. Such an approach does not allow one to trace the dynamics of behavior of a number of informative characteristics (such as the absorption line shape or the population of the Zeeman sublevels) that depend on parameters being varied (intensities of the orienting and test fields, branching factor, etc.). It cannot be ruled out that a number of important features of the processes under investigation will be lost since it is difficult to account for these processes (e.g., the effect of light pressure) correctly. Modern computational methods remove these limitations to a considerable extent and make it possible to analyze the kinetic equation for arbitrary values of intensities of the orienting and test fields as well as the total angular momenta of the lower and upper states of the atom.

In this study, we report on the results of investigations of the shape of the absorption spectrum of the test field in the presence of a strong orienting field on transition in the Ne atom ($J = 2 \longrightarrow J = 2$, $J = 2 \longrightarrow J = 1$). The results were obtained by solving the equations numerically in a large range of atomic system parameters (branching factors), polarization, and intensity of the strong field. Calculations are based on the analysis of resonant interaction of a high-intensity polarized orienting field with degenerate metastable atomic systems, which was developed in [4]. It was found that the

shape of the saturated absorption resonance near the line center depends on the branching factor and the dipole moment of the transition between magnetic sublevels. The resonance has a Doppler contour with a traditional Lamb-type dip and with an absorption peak, whose amplitudes and widths are determined by the intensity (saturation parameter) of the strong field. On the transition $J = 2 \longrightarrow J = 2$, a doublet structure appears in the shape of the resonance absorption peak near the line center for certain values of the branching factor and the saturation parameter, while no doublet splitting in the resonance peak is observed on the transition $J = 2 \longrightarrow J = 1$. The second part of this article is devoted to analysis of the effect of the light pressure force on the shape of the saturated absorption resonance. The results of numerical analysis are used for explaining the experimental data obtained for the Ne atom on the transition $1s_5 \longrightarrow 2p_8$ and partly reported in [3].

2. THEORETICAL MODEL

The problem of interaction of metastable states of atoms with laser radiation is essentially nonlinear and complicated in view of the specific features of such states, and requires specification of the energy level diagram and the properties of resonant radiation. For this reason, we will consider transitions between the *m*th and the *n*th levels of the Ne atom with total momenta $J_n = 2 \longrightarrow J_m = 2$ and $J_n = 2 \longrightarrow J_m = 1$. However, the results are also valid for transitions in other atoms with the same values of total angular momenta.

We will consider the problem of the absorption spectrum of a test field in the presence of counterpropagating radiation from a strong field of the same frequency. Radiation from a strong field of strength $\mathbf{E}_1$ is assumed to have the form of a plane monochromatic wave (of frequency ω and wave vector $\boldsymbol{\kappa}$), which is in resonance with the *m*–*n* atomic transition (transition frequency is ω_{mn}) with a linear or circular σ^+ polarization.

Radiation emitted by a test field of strength $\mathbf{E}_2$ is presented by a plane monochromatic wave (frequency $\omega_2 = \omega$ and wave vector $\boldsymbol{\kappa}_2 = -\boldsymbol{\kappa}$) with a circular σ^- polarization, which propagates in a direction opposite to that of the high-intensity light wave. Such a formulation of the problem is analogous to the experimental conditions in [3]. The case of a test wave with a linear polarization orthogonal to the strong field can also be reduced to the problem in question.

We assume that the gas pressure is quite low; in this case, collisions can be neglected and the only relaxation mechanism is associated with spontaneous transitions in the atom.

In the case of the linear polarization of the strong field, we will consider the problem in a coordinate system with the quantization axis *z* directed along the field $\mathbf{E}_1$. In this coordinate system, the strong field induces transitions involving a change in the magnetic quantum

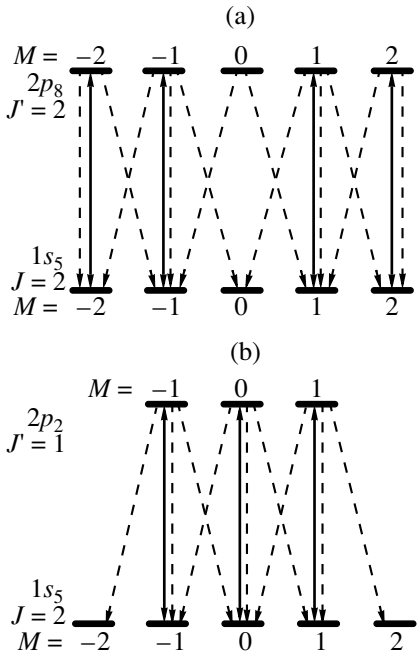


Fig. 1. Diagram of spontaneous and induced transitions between the degenerate upper (*m*) and lower (*n*) levels of the Ne atom: (a) $J_n = 2 \longrightarrow J_m = 2$; (b) $J_n = 2 \longrightarrow J_m = 1$.

number $\Delta M = 0$ (Fig. 1); spontaneous transitions in this case are observed with a change of $\Delta M = 0, \pm 1$. In the given problem, we assume that the strong field sets a nonequilibrium population of magnetic sublevels and the absorption coefficient for the weak field is determined by this nonequilibrium difference in the populations.

We will solve this problem proceeding from the kinetic equations for the density matrix of the atomic system, which, according to [4, 6] and the notation adopted there, have the following form in steady-state conditions:

$$\Gamma_n \rho_M = -W_M(\rho_M - r_{M1}) + \sum_{M'} A_{MM'} r_{M'} + Q_M, \quad (1)$$

$$\Gamma_m r_{M1} = W_M(\rho_M - r_{M1}) + q_{M1}. \quad (2)$$

Here, ρ_M and r_M are the diagonal elements of the density matrix describing the population of the lower and upper magnetic sublevels, $|M|, |M'| \le 2$;

$$W_M = \frac{2\Gamma |G_M|^2}{\Gamma^2 + \Omega_1^2}$$

is the probability of the induced transitions $nM \longleftrightarrow mM1$; $\Omega_1 = \Omega - \boldsymbol{\kappa} \cdot \mathbf{v}$, $\Omega = \omega - \omega_{mn}$ being the detuning of the strong field frequency from the frequency of the

resonant transition of the atom with allowance for the Doppler shift; $\Gamma = (\Gamma_m + \Gamma_n)/2$ is the transition width;

$$G_M = \frac{dE}{2\sqrt{3}\hbar}(-1)^{J_n - M}\langle J_m M J_n - M | 10\rangle$$

is the parameter of the interaction of the atom with the strong field (in the case of a strong field with a circular polarization, $G_M^+ = G_M/\sqrt{2}$); Γ_m, Γ_n, q_M, and Q_M are the rates of relaxation and excitation of sublevels mM and nM; $\mathbf{v}$ is the velocity of the atom; d is the reduced dipole moment of the m–n transition; and the expression in the angle brackets is the coefficient of vector summation. The rates of the spontaneous transitions mM'–nM are determined by the Einstein coefficients A_{mn} and the vector summation coefficients. In Eqs. (1) and (2), the subscripts $M1 = M$ in the case of linear polarization of the strong field and $M1 = M + 1$ in the case of circular polarization.

The absorption coefficient of the counterpropagating test light wave with circular polarization (all subsequent results are given for the σ^- polarization) per atom was defined as

$$K_s^- = K_0 \sum_{-2 \le M \le 2} |d_{MM-1}|^2 \times \left\langle \frac{\Gamma^2}{\Gamma^2 + (\Omega + \boldsymbol{\kappa}\cdot\mathbf{v})^2}(\rho_M - r_{M-1}) \right\rangle, \quad (3)$$

where

$$K_0 = \frac{4\pi\omega_{mn}}{c\hbar\Gamma},$$

d_{MM-1} is the dipole moment of the transition between the magnetic sublevels, and the angle brackets indicate averaging over velocities. Averaging in expression (3) was carried out with equilibrium (Maxwellian) and nonequilibrium velocity distributions of particles.

System of equations (1), (2) together with Eq. (3) was solved numerically on a mesh with a step of relative frequency detuning of $\Omega/\boldsymbol{\kappa}\cdot\mathbf{v}_t$ and a relative particle velocity v/v_t of 10^{-3}, where v_t is the most probable velocity of the ensemble. In averaging, particles with velocities $|v/v_t| \le 3$ were taken into account. It can be seen that the number of discarded particles is exponentially small. The following values of the atomic transition parameters were taken for calculations [5]: the probabilities of spontaneous transitions with $J = 2 \longrightarrow J = 2$ and $J = 2 \longrightarrow J = 1$ were $A_{mn} = 1.36 \times 10^7\ \mathrm{s}^{-1}$ and $A_{mn} = 3.34 \times 10^7\ \mathrm{s}^{-1}$, respectively; the rate of decay from the lower level was $\Gamma_n = 10^5\ \mathrm{s}^{-1}$; and the rate of decay from the upper level, Γ_m, was determined in terms of the branching parameter $\alpha = A_{mn}/\Gamma_m$, the value of α being varied in the range 0.1–1. The strong field saturation parameter, defined as

$$\Re = \frac{2|G_M|^2}{\Gamma\Gamma_n},$$

was varied in the range 0.5–10^4. Excitation rates Q_M and q_M determine the population of states of the Ne atom in the absence of a strong field. Since the population n of the lower state in a gas discharge is considerably higher than the population m of the upper state, we set $q_M = 0$ in numerical calculations and assumed that the pumping of the lower state is the same to all sublevels with a rate $Q_M = \Gamma_n$, although the population of sublevels with $|M| = 2$ under the conditions of the discharge cell is slightly higher than the population of the remaining sublevels [7].

3. SINGULARITIES IN THE ABSORPTION SPECTRUM OF THE TEST FIELD ON THE TRANSITION $J_n = 2 \longrightarrow J_m = 2$ OF THE Ne ATOM

Transition $J_m = 2 \longrightarrow J_n = 2$ is characterized by the following features: the dipole moment and, hence, the probability of a spontaneous transition between magnetic sublevels with $M = 0$ are equal to zero ($d_{00} = 0$, $A_{00} = 0$), while the ratio of the maximal probability A_{22} to the minimal probability A_{11} is equal to 4. This leads to an elevated population n of the level with $M = 0$ (in view of the absence of an induced transition from this sublevel in a strong field; see Fig. 1a) and to significantly different effects (broadening) of the strong field on the shape of the absorption line of the weak field on transitions between individual magnetic sublevels. This is manifested in the complex form of the spectrum of the resultant absorption coefficient in the vicinity of the center of the transition line.

Numerical investigations in the given model problem show that the absorption spectrum of the test field on a transition between degenerate states of the Ne atom with total angular momenta $J_n = J_m = 2$ is determined by the result of summation of four (and not two, as indicated in [4]) Lorentz-type contours against the Doppler background. These contours are formed as a result of absorption from the corresponding magnetic sublevels of the lower state and have different amplitudes and widths near the transition line center due to different values of the dipole moments of the transitions between the magnetic sublevels and of their occupancies under the action of a strong light field.

It was found that the action of a strong field leads to the following dynamics in the velocity distribution of particles on the magnetic sublevels of the lower state:

(a) the formation of a dip in the particle distributions function (Bennett hole) on sublevels with $M = \pm 2$ for any values of saturation parameter $\Re$;

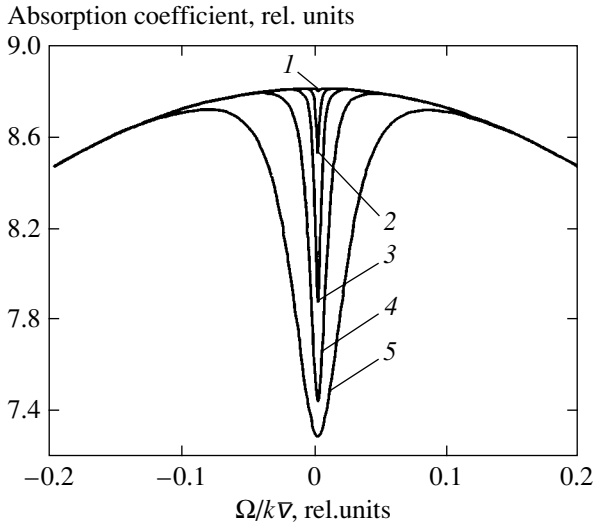


Fig. 2. Shape of the test field absorption line calculated taking into account the equilibrium velocity distribution of particles for branching parameter $\alpha = 0.8$ and various values of strong field saturation parameter $\Re = 1$ (*1*), 10 (*2*), 10^2 (*3*), 10^3 (*4*), and 10^4 (*5*).

(b) the emergence of a small-amplitude peak on sublevels with $M = \pm 1$ for $\Re \leq 1$ and of a dip with a width much smaller than on sublevels with $M = \pm 2$ for $\Re > 1$;

(c) the formation of a peak on a sublevel with $M = 0$ for any saturation parameter $\Re$. It is the result of summation of these contours with different amplitudes and widths that determines the type of the behavior of the saturation absorption resonance shape, including the doublet structure in the resonance peak discovered in experiment [3].

Numerical analysis reveals that the shape of the absorption line of the test field is determined by the branching parameter of the atomic system, which is defined as $\alpha = A_{mn}/\Gamma_m$. For values of $\alpha < 0.85$, the absorption spectrum near the center of the transition line is a conventional saturated absorption resonance in the form of a Lorentzian dip against the background of a broad Doppler absorption contour with a width and an amplitude depending on the intensity (saturation parameter) of the strong field (Fig. 2).

In the range $0.85 \leq \alpha \leq 1$, an absorption peak with an amplitude and a spectral width depending on the intensity of the strong field is formed instead of the dip. In the vicinity of $\alpha \approx 0.85$, a complex field dependence of the saturated absorption resonance spectrum on the saturation parameter of the strong field is observed (Fig. 3).

For values of saturation parameter $\Re \lesssim 1$, a small-amplitude peak is observed (curve *1* in Fig. 3). With increasing intensity of the strong field, for a saturation parameter of $\Re \lesssim 10$, this peak splits into two components (characteristic curve *2*), while for values of $\Re \geq 10$, a dip with a width and an amplitude determined by the intensity of the strong field is formed (curves *3*–*6*). In this case, the maximum amplitude of the dip considerably exceeds the amplitude of the peak.

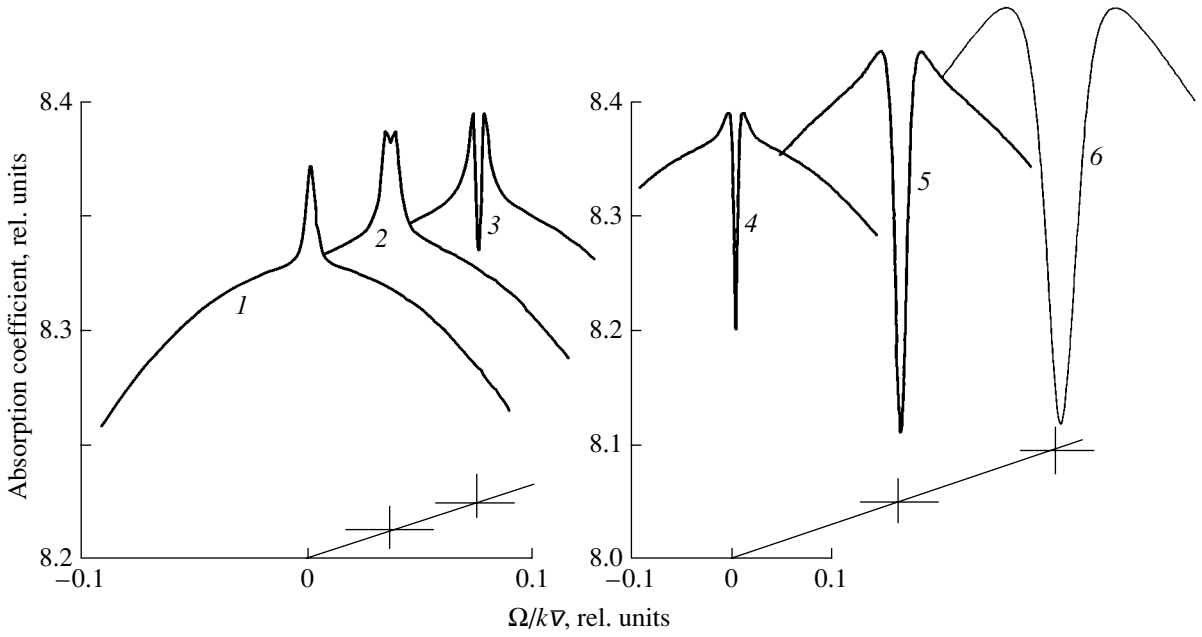


Fig. 3. Shape of the test field absorption line calculated with allowance for the equilibrium velocity distribution of particles for branching parameter $\alpha = 0.85$ and various values of strong field saturation parameter $\Re = 1$ (*1*), 5 (*2*), 10 (*3*), 10^2 (*4*), 10^3 (*5*), and 10^4 (*6*).

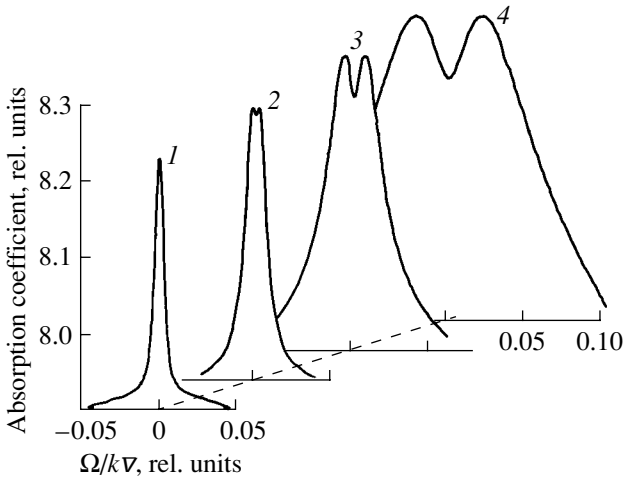


Fig. 4. Shape of the peak of the test field absorption line calculated for various values of strong field saturation parameter $\Re = 50$ (*1*), 10^2 (*2*), 10^3 (*3*), and 10^4 (*4*) with allowance for the equilibrium velocity distribution of particles for branching parameter $\alpha = 0.9$.

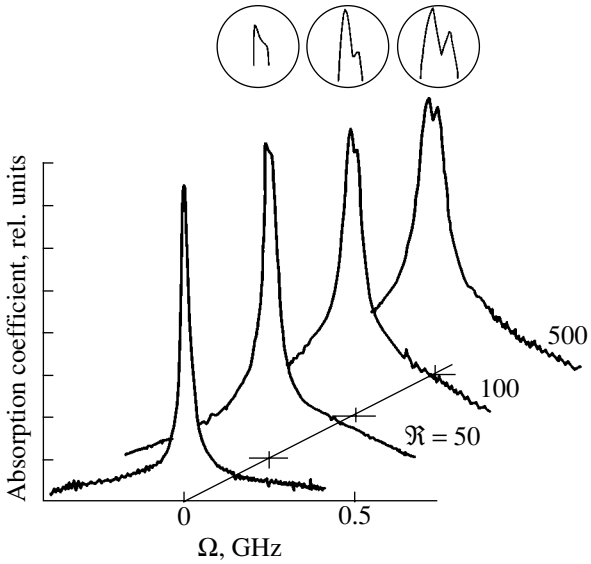


Fig. 5. Experimental shape of the test field absorption line for various values of the strong field saturation parameter.

In the range of the branching parameter $0.9 \le \alpha \le 0.95$ and for saturations parameters $\Re \ge 50$, the resonance peak splits into two components. The dependence of the peak amplitude of the strong field intensity exhibits saturation, while the peak width and the value of the doublet splitting strongly depend on the field intensity in the range of values under investigation. Figure 4 shows typical contours of the absorption line of the test field as functions of the strong field absorption parameter for branching parameter $\alpha = 0.9$.

Figure 5 shows the experimental frequency dependences of the absorption coefficient of the test field on the saturation parameter of a high-intensity light wave from [5]. A comparison of the shape of the curves in Figs. 4 and 5 shows that the experimental and calculated data are in good quantitative agreement. It follows from the results of calculations that the variation of the saturation parameter of strong field in the range of $\Re = 10^2$–10^4 leads to the variation of the half-amplitude width of the absorption resonance peak in the range 12–95 MHz, while the doublet splitting varies in this case from 3 to 33 MHz. According to experimental data, the maximal value of frequency between peaks of the doublet splitting was about 25 MHz for a value of parameter $\Re \sim 500$, while the absorption peak width varied from 14 to 70 MHz upon a change in the saturation parameter in the range 100–500.

It should be emphasized that model calculations lead to a symmetric splitting of the peak of the saturated absorption resonance line, while asymmetric splitting of the resonance line was observed in experiments [5]. It should be noted in this connection that the value of the branching parameter obtained from the data [4] on the rates of spontaneous decay channels of the $2p_8$ level of the Ne atom is $\alpha = 0.336$. In accordance with the results described above, the absorption spectrum of the test field must have a dip for this value of α (see Fig. 2). However, in experiments, an absorption peak was observed, indicating the emergence (under the given experimental conditions [5]) of physical processes that considerably increase the contribution from the transition on which the strong field is acting and reduces the contribution from other relaxation channels of the upper level. The nature of these processes still remains unclear.

For values of $\alpha > 0.95$, the doublet structure at the line center is not manifested and the resultant spectrum has the form of an absorption peak (Fig. 6) whose amplitude and width are determined by the intensity of the strong field.

A similar analysis for the case of a strong field with circular polarization, inducing transition with $\Delta M = \pm 1$, leads to a qualitatively similar behavior of the saturated absorption resonance of the test light wave, the only difference being in the values of the peak width and the doublet splitting.

Concluding this section, let us consider the physical origin of the strong dependence of the shape of the absorption line for the test field on the branching coefficient α. This coefficient essentially determines the fraction of the "useful" decay rate of level *m* via channel *m*–*n* of the total decay rate of the level to all the states optically coupled with the given level. In this connection, the values of parameter α are determined by the amplitudes of resonances as well as by their widths in the spectroscopy of coupled transitions [6]. This effect is manifested most clearly in the case of degenerate atomic systems with a long-lived lower level, when the strong optical pumping field and subsequent spontaneous transitions couple a considerable

number of two-level systems. This leads to the decisive role of the optical pumping effect in the population of a number of magnetic sublevels of the lower state.

It was noted above that transition $J_n = 2 \longrightarrow J_m = 2$ in the Ne atom is characterized by different widths of the Bennett structures in the populations of the magnetic sublevels of the lower and upper states of the transition, which are coupled by a strong pumping field, for saturation parameters $\Re \gg 1$ (see Fig. 1a). In this case, the population of the sublevel with $M = 0$ of the lower state is determined by the number of cascade-type spontaneous transitions via the upper state; the contribution to its population from other magnetic sublevels is proportional to $\alpha^{|M|}$.

The contour of the absorption line of the test field is the result of summation of the four contours corresponding to transitions between individual magnetic sublevels. For this reason, a competition appears between the dips and Bennett peaks in the differences in the populations of the sublevels of the lower and upper states with $M = \pm 2, \pm 1$, which are coupled by the strong pumping field, as well as the Bennett peak in the population of the lower state with $M = 0$.

For small values of α ($\alpha \leq 0.8$), the fraction of particles reaching the sublevel with $M = 0$ from the lower state as a result of optical pumping is relatively small (particles are mainly pumped to the third levels via the channel $n \longrightarrow m \longrightarrow j \neq n$). For this reason, the resultant contour of the test field absorption line is determined by the Bennett structures in the population of the energy levels coupled by the strong field and has the shape of a Doppler-type absorption line with a traditional dip at the line center (see Fig. 2).

With increasing α, the fraction of particles reaching the sublevel with $M = 0$ as a result of optical pumping increases, which leads to an increase in the contribution of the given sublevel to the shape of the absorption coefficient of the test field in the form of a peak at the line center (for $\alpha \sim 1$, this contribution becomes decisive; Fig. 6). In a certain range of α values (0.85–0.9), the amplitudes of dips and peaks of the Bennett structures are almost identical, while their widths differ substantially. This leads to a complex dependence of the total absorption coefficient of the test field. Figures 3 and 4 reflect the dynamics of the saturated absorption resonance spectrum for a varying intensity of the strong field in the given range of α.

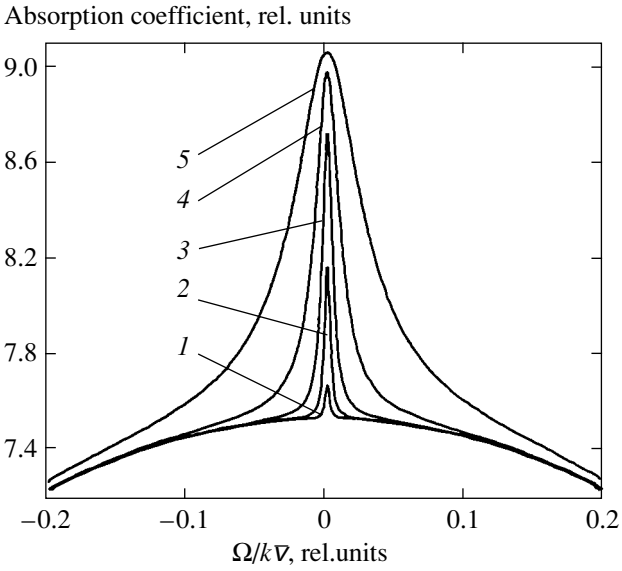


Fig. 6. Shape of the test field absorption line calculated with allowance for the equilibrium velocity distribution of particles for branching parameter $\alpha = 0.98$ and various values of strong field saturation parameter $\Re = 1$ (*1*), 10 (*2*), 10^2 (*3*), 10^3 (*4*), and 10^4 (*5*).

4. SINGULARITIES IN THE TEST FIELD ABSORPTION SPECTRUM IN THE TRANSITION $J_n = 2 \longrightarrow J_m = 1$ IN THE Ne ATOM

In the case of an atomic transition with $J_n = 2 \longrightarrow J_m = 1$, the dipole moments and the probabilities of spontaneous transitions between magnetic sublevels differ insignificantly (the values of d_{00} and A_{00} differing from zero), while the ratio of the maximal probability A_{00} to the minimal probability A_{11} ($A_{11} = A_{-1-1}$) is equal to 4/3. In the given case, the strong field $\mathbf{E}_1$ with linear polarization, inducing transitions without a change in the magnetic quantum number ($\Delta M = 0$), leads to a decrease in the population of the lower levels with $M = 0, \pm 1$ as a result of optical pumping (the particle distribution function acquires dips) and to an increase in the population of levels with $M = \pm 2$ (a peak appears in the distribution function), the difference in the field-induced broadening of the peaks and the dips being insignificant. These circumstances make the situation substantially different from that considered above for the transition between the energy levels with the total angular moments $J_n = J_m = 2$.

The numerical solutions of the modified system of equations (1)–(3) with the corresponding values of transition parameters (spontaneous transition probabilities, dipole moment, and vector summation coefficient) give the following singularities in the form of the absorption coefficient of the antiparallel test field with circular polarization.

1. The absorption spectrum of the test field in the vicinity of the center of the atomic transition line is the sum of three Lorentzian profiles against a broad Doppler background line. In this case, a dip is observed in the shape of the line near its center on transitions from magnetic sublevels with $M = 0, \pm 1$, while a peak appears on transitions from sublevels with $M = \pm 2$. It is important that the amplitudes and spectral widths of the peak and the dips differ insignificantly in view of small differences in the probabilities and dipole moments for transitions from different magnetic sublevels.

2. The resultant absorption spectrum also depends on branching coefficient α: for values of $\alpha < 0.65$ a characteristic dip is observed near the line center, while the absorption peak is observed instead of the traditional dip for $\alpha = 0.65$–1. The spectral widths of the peak and the dip are determined by the intensity of the strong field, while their frequency behavior is similar to the dependences depicted in Figs. 2 and 6. In this case, the doublet splitting structure typical of the absorption peak on the transition $J_n = 2 \longrightarrow J_m = 2$ does not appear in the absorption peak.

5. EFFECT OF LIGHT PRESSURE ON THE TRANSITION $J_n = 2 \longrightarrow J_m = 2$ IN THE Ne ATOM

The model calculations with an equilibrium particle distribution over magnetic sublevels led to symmetric resonance peak splitting (see Fig. 4), while the experimentally observed amplitudes of the peak components are different (see Fig. 5). In this connection we analyzed the effect of the force of light pressure (associated with the strong field) on the shape of the saturated absorption resonance. The analysis was based on the technique developed by us earlier [8, 9].

In contrast to transitions from the ground state, the action of the force of light pressure for transitions from excited states of atoms exhibit features associated with the following factors.

(a) A finite lifetime of the lower state and the existence of several decay channels for the upper state. In this case, time t_r of the resonant interaction of an atom with the strong field is determined by the spontaneous transition probability and the branching parameter α ($t_r = A_{mn}^{-1}/(1 - \alpha)$). Time t_r is found to be the same for all atoms in the ensemble in contrast to the case of interaction between atoms in the ground state, while the time of the interaction is determined by the time of flight of the atom through the light beam.

(b) A difference in the values of the dipole moment and probabilities of transitions between degenerate sublevels with different values of magnetic quantum number M. As a result, the light pressure due to the strong field produces different effects on the particle distribution on these magnetic sublevels. Analysis shows that the maximum effect of the strong field is manifested in the particle distribution over sublevels with $M = \pm 2$.

The particle distribution over sublevels in the field of a high-intensity light wave was determined from the solution of a Fokker–Planck equation having the form

$$\frac{\partial f(v_z, t)}{\partial t} + \frac{\partial (A_z f(v_z, t))}{\partial v_z} + \frac{\partial^2 (C_{zz} f(v_z, t))}{\partial v_z^2} = 0 \quad (4)$$

for a small Doppler shift of the atomic frequency; $\boldsymbol{\kappa} \cdot \mathbf{v}_r = \hbar \boldsymbol{\kappa}^2 / m_a$ (m_a is the atomic mass) versus the uniform transition width ($\boldsymbol{\kappa} \cdot \mathbf{v}_r \ll \Gamma$) [10, 11]. Here, factor A_z is determined by the light pressure force,

$$A_z = \pm \frac{\Gamma v_z \Re}{[1 + \Re + (\Omega - \kappa \upsilon_z)^2/\Gamma^2]}, \quad (5)$$

while the quantity C_{zz} determining particle diffusion in the velocity space has the form

$$C_{zz} = \frac{0.5 \Gamma v_r^2 \Re}{1 + \Re + (\Omega - \kappa v_z)^2/\Gamma^2}. \quad (6)$$

The plus and minus signs correspond to particles moving parallel and antiparallel to the wave vector $\boldsymbol{\kappa}$ of the strong light field.

Fokker–Planck equation (4) was solved numerically on a 2D 600×1000 mesh with a step of relative velocity variation of $\kappa \Delta v_z/\Gamma = 0.5$–0.1 and a step of relative time variation ($\tau_r = \kappa v_r t$) $\Delta\tau = 0.1$–0.05. In our calculations, we used the following values of atomic parameters: the most probable velocity $v_t \approx 7 \times 10^2 \Gamma/\kappa$ at a gas temperature of $T = 300$ K, a radiation wavelength $\lambda \approx 633$ nm, a recoil frequency $\Delta\omega_r = \kappa v_r = 3.3 \times 10^5$ s^{-1}, a ratio $\Delta\omega_r/\Gamma \sim 5 \times 10^{-2}$, and values of saturation parameter $\Re$ varied in the range 0.1–10^4. The relative time of interaction of an atom with the strong field (during which the particle velocity distribution function changes) was $\tau_r \approx 0.5$–1 for values of branching parameter $\alpha \sim 0.9$–0.95. It should be recalled that the doublet form of the saturated absorption spectrum is observed for the given values of the parameters.

The results of calculations of the behavior of the saturated absorption resonance shape near the center of the transition line of the Ne atom depending on the strong field saturation parameter $\Re$ taking into account the nonuniform distribution of particles are shown in Fig. 7.

The behavior of the curves demonstrates that the action of light pressure force is manifested in the asymmetry of the absorption peak for $\Re \le 50$ and in the amplitude ratio of the peak splitting doublet for $\Re \ge 100$.

Figure 8 shows the frequency behavior of the relative change in the test field absorption coefficient, $(K_0^- - K_s^-)/K_0^-$ for various intensities of the strong field, directly reflecting the manifestation of the effect of light pressure on the transition being studied. It can be seen that the experimental [5] and calculated data on the relative change in the amplitudes of the splitting doublet are in good quantitative agreement: the maxi-

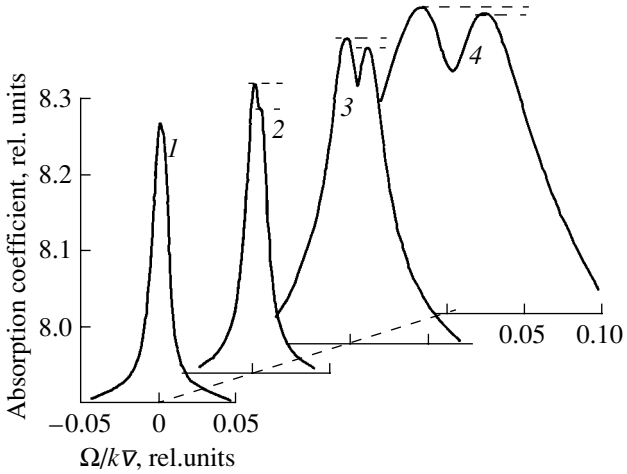


Fig. 7. Shape of the peak of the test field absorption line calculated for various values of strong field saturation parameter $\Re = 50$ (*1*), 10^2 (*2*), 10^3 (*3*), and 10^4 (*4*) with allowance for the effect of light pressure for branching parameter $\alpha = 0.9$.

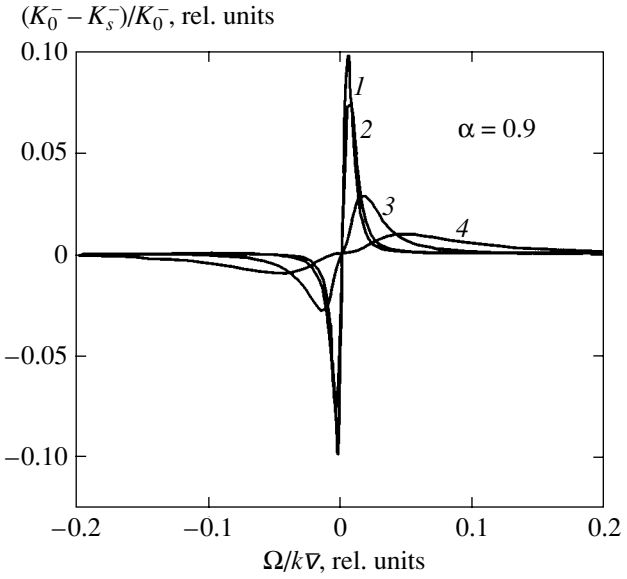


Fig. 8. Dependence of the relative variation of the test field absorption coefficient for various values of strong field saturation parameter $\Re = 50$ (*1*), 10^2 (*2*), 10^3 (*3*), and 10^4 (*4*).

mal calculated value amounts to about 9%, while the experimental value is approximately equal to 8%.

It follows from [8] that the effect of light pressure on the spectral parameters of the saturated absorption resonance (the shape of the resonance and the position of its minimum relative to the atomic transition frequency) is maximal for values of the saturation parameter $\Re = 1$–2. However, the peculiarity of the formation of the doublet spectral structure on transitions with degenerate states of the atom as a result of subtraction of several contours (realization of a difference scheme of observation on atomic level) makes it very sensitive to the distribution function of particles interacting with the field and permits the observation of light pressure force in the region nonoptimal for the effect.

6. CONCLUSIONS

Thus, the results described above show that, with increasing degeneracy of the levels participating in the resonant interaction of an atom with optical fields and the lifetime of these levels, the potentialities of saturated absorption spectroscopy as regards obtaining narrow resonance structures considerably increase. This is manifested most clearly under the optical pumping conditions in high-intensity laser fields. It should be noted that resonances anomalous in widths and amplitude signs were observed even in the first experiments with high intensities of orienting fields [1–3]. The reasons for such anomalies have not yet been unambiguously interpreted. For example, under our experimental conditions, the mechanism of the effective increase in branching parameter α of the Ne atom in the discharge conditions for the $1s_5$ level from the block of closely spaced $1s_i$ levels remains unclear. A possible reason for this increase can be atomic collisions with electrons, leading to mixing of populations of these levels in the block of $1s_i$ levels.

In addition, the above results indicate that light pressure has a considerable effect on the parameters of the saturated absorption resonance on transitions from excited states. The possibility of manifestations of the light pressure effect on the spectroscopic parameters of transitions in excited atoms were not considered earlier. The intensity of the effect can be significant, especially for light atoms. The latter circumstance appears as especially important for interpreting the results of precision metrological measurements on the basis of the saturated absorption method, including the Rydberg constant [12].

ACKNOWLEDGMENTS

We are grateful to S.G. Rautian and É.G. Saprykin for fruitful discussions of the results.

This study was party supported by the program Russian Universities (grant no. UR.01.01.031) and the Russian Foundation for Basic Research (project no. 04-02-17552).

REFERENCES

1. W. Gawlik and G. W. Series, in *Proceedings of 4th International Conference on Laser Spectroscopy, Rottach-Egern, Germany, 1979*, Ed. by H. Walter and M. Rothe (Springer, Berlin, 1979), Vol. 21, p. 210.

2. D. E. Murnick, M. S. Feld, M. M. Burns, *et al.*, in *Proceedings of 4th International Conference on Laser Spectroscopy, Rottach-Egern, Germany, 1979*, Ed. by H. Walter and M. Rothe (Springer, Berlin, 1979), Vol. 21, p. 195.

3. I. A. Kartashov and A. V. Shishaev, Pis'ma Zh. Éksp. Teor. Fiz. **58**, 501 (1993) [JETP Lett. **58**, 502 (1993)].

4. S. G. Rautian and A. V. Shishaev, Zh. Éksp. Teor. Fiz. **108**, 807 (1995) [JETP **81**, 440 (1995)].

5. I. A. Kartashov, S. G. Rautian, and A. V. Shishaev, in *Physics of Vibrations* (Allerton, New York, 1998), Vol. 6, p. 143.

6. S. G. Rautian, G. I. Smirnov, and A. M. Shalagin, *Nonlinear Resonances in Atomic and Molecular Spectra* (Nauka, Novosibirsk, 1979) [in Russian].

7. E. B. Aleksandrov, G. I. Khvostenko, and M. P. Chaĭka, *Interference of Atomic States* (Nauka, Moscow, 1991) [in Russian].

8. A. A. Chernenko and A. V. Shishaev, Opt. Spektrosk. **93**, 401 (2002) [Opt. Spectrosc. **93**, 368 (2002)].

9. A. A. Chernenko and A. V. Shishaev, Opt. Commun. **211**, 249 (2002).

10. V. G. Minogin and V. S. Letokhov, *The Pressure of Laser Radiation on Atoms* (Nauka, Moscow, 1986) [in Russian].

11. A. P. Kazantsev, G. I. Surdutovich, and V. P. Yakovlev, *The Mechanical Action of Light on Atoms* (Nauka, Moscow, 1991) [in Russian].

12. C. Wieman and T. W. Hansch, Phys. Rev. A **22**, 192 (1980).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 99, No. 2, 2004, pp. 279–285.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 2, 2004, pp. 315–322.

ATOMS, SPECTRA, RADIATION

Internal Modes of Solitons in a Bose–Einstein Condensate

P. I. Krepostnov, V. O. Popov, and N. N. Rozanov

Research Institute for Laser Physics, Vavilov State Optical Institute, St. Petersburg, 199034 Russia
e-mail: nrosanov@yahoo.com; rosanov@ilph.spb.su

Received November 20, 2003

Abstract—A modified Gross–Pitaevskii equation for a Bose–Einstein condensate with nonlocal interaction between atoms is used to analyze and calculate the characteristics of internal modes of bright solitons (eigenmodes of small perturbations of the condensate). The spectra of even and odd internal modes are found for one- and two-dimensional solitons, and the rate constant characterizing nonlinear internal-mode damping is determined. The possibility of experimental verification of the results is discussed. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Analysis of small-amplitude collective oscillations of an atomic Bose–Einstein condensate (BEC) confined in a trap provides an effective tool for analyzing this macroscopic quantum object and verifying its theoretical description [1–6]. Recent experimental observations of bright solitons in lithium BECs [7, 8], which are characterized by a negative scattering length, have demonstrated the possibility of studying an untrapped condensate. This offers new opportunities for analysis of the physics of BECs, including analysis of the eigenmodes of small-amplitude perturbations of BEC solitons (usually called internal modes).

In the standard approximate description of BECs based on the Gross–Pitaevskii equation [1], interaction between atoms is assumed to be local (the interatomic potential is reduced to the delta function of the interatomic distance). Under this assumption, stable nonlinear localization (due to interatomic interaction) is possible only along one direction. In other words, only quasi-one-dimensional BEC solitons can be stable, provided that the condensate can be localized along the remaining two coordinates by applying an external field. This scheme was implemented in the experiment reported in [7, 8].

It was shown in [9] for several model examples that collapse can be prevented by using a nonlocal potential, which implies that two- and three-dimensional BEC solitons can be obtained. An analysis of the Gross–Pitaevskii equation modified to describe weakly nonlocal interatomic interaction predicts that sufficiently large two-dimensional BEC solitons with negative scattering lengths are stable and that internal modes of these solitons exist [10]. In [11, 12], one-, two-, and three-dimensional BEC solitons were obtained as numerical solutions to the modified Gross–Pitaevskii equation and it was pointed out that solitons of this kind can be used in creating nanostructures and in nanolithography.

In this paper, we analyze the effect of nonlocality of interatomic interaction on the properties of one-dimensional BEC solitons. Stable solitons can be obtained without assuming nonlocality, but they do not have any internal modes. In Section 2, we show that nonlocal interaction gives rise to an internal-mode spectrum when a certain threshold is exceeded and determine some characteristics of even and odd internal modes. The internal-mode oscillations are not damped in the linear approximation. In Section 3, we apply the semianalytical approach developed in [13–16] and validated numerically in [17] to analyze the rate of nonlinear damping due to outward flow of atoms (which escape to infinity since there is no trap). Section 4 presents analogous results concerning two-dimensional BEC solitons, a discussion, and conclusions.

2. INTERNAL MODES OF A ONE-DIMENSIONAL BOSE–EINSTEIN CONDENSATE SOLITON

In the case of a negative scattering length, the modified Gross–Pitaevskii equation for a single-particle wave function $\Phi(x, t)$ of a system with a weakly nonlocal interatomic interaction is written as

$$i\frac{\partial\Phi}{\partial t} + \frac{\partial^2\Phi}{\partial x^2} + \Phi|\Phi|^2 + \Phi\frac{\partial^2|\Phi|^2}{\partial x^2} = 0, \qquad (1)$$

where t and x are dimensionless time and coordinate, respectively, and all parameters are eliminated by appropriate normalization [11, 12]. Note that renormalization can be used to ensure that Eq. (1) is applicable

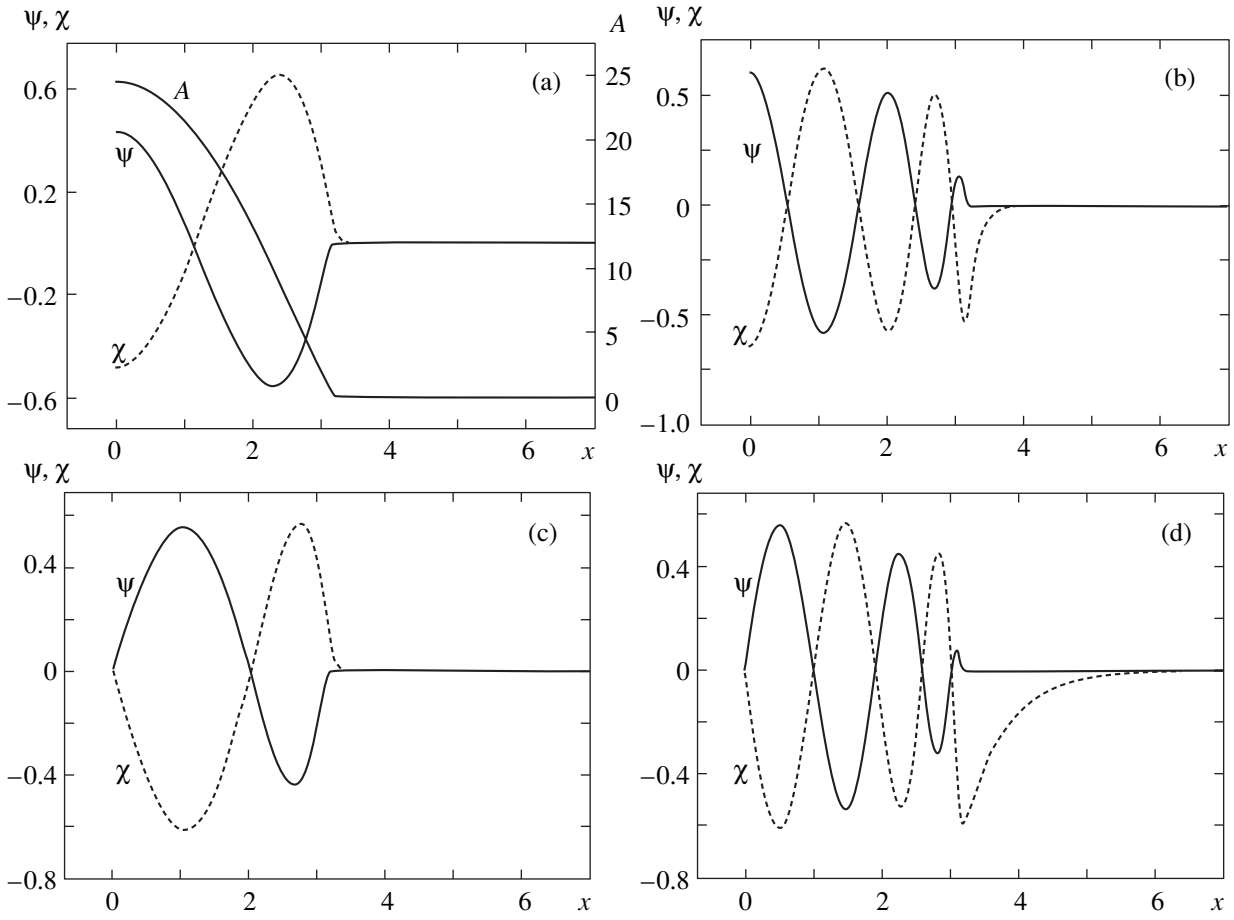


Fig. 1. One-dimensional BEC soliton with $\kappa = 300$: (a) amplitude $A(x)$ and eigenfunctions $\psi(x)$ and $\chi(x)$ of (a) the first even, (b) the fourth even, (c) the first odd, and (d) the fourth odd internal modes.

when the divergence of the integrals representing the moments of the interatomic potential is eliminated by a more general method than that proposed in [11].

Stationary solutions to (1) have the form

$$\Phi(x, t) = A(x)\exp(i\kappa t), \tag{2}$$

where the function $A(x)$ (exponentially decreasing as $x \longrightarrow \pm\infty$) is an eigensolution of the real nonlinear ordinary differential equation

$$\frac{d^2A}{dx^2} - \kappa A + A^3 + A\frac{d^2A^2}{dx^2} = 0. \tag{3}$$

The characteristics of one-, two-, and three-dimensional solitons were determined by solving (3) numerically in [12]. Equation (3) can be solved analytically by changing to the variable $p = dA/dx$. The solutions to the resulting first-order linear differential equation in p^2 can be expressed in terms of elliptic integrals of the first and third kinds in the general case and in terms of elementary functions in some special cases [18]. For the localized symmetric solution of interest here, A and x are related as follows (see Fig. 1a):

$$|x| = \int_A^{A_m} \sqrt{\frac{A^2 + \frac{1}{2}}{\frac{1}{2}\kappa A^2 - \frac{1}{4}A^4}}\,dA = 2\operatorname{arccot} q + \frac{1}{2\sqrt{\kappa}}\ln\left|\frac{2q\sqrt{\kappa}+1}{2q\sqrt{\kappa}-1}\right|, \quad q = \sqrt{\frac{A^2+\frac{1}{2}}{2\kappa - A^2}}, \tag{4}$$

where $A_m = \sqrt{2\kappa}$ is the peak soliton amplitude (at $x = 0$). As $x \longrightarrow \infty$, solutions have the asymptotic form

$$A(x) \sim \exp(-\sqrt{\kappa}|x|).$$

The spectral parameter characterizing a soliton, $\kappa > 0$, is proportional to the chemical potential (energy) of the condensate and is uniquely related to the following

quantity proportional to the number of particles (see Fig. 2):

$$N = \int_{-\infty}^{\infty} A^2(x)dx = 2\sqrt{\kappa} + (4\kappa + 1)\operatorname{arccot}\frac{1}{2\sqrt{\kappa}}. \quad (5)$$

If $\kappa \ll 1$, then $N \approx 4\sqrt{\kappa}$; if $\kappa \gg 1$, then $N \approx 2\pi\kappa$.

Let us introduce a small deviation of the wave function from (2):

$$\Phi(x, t) = [A(x) + \delta\Phi^{(1)}(x, t)]\exp(i\kappa t). \quad (6)$$

The governing equation (1) linearized with respect to $\delta\Phi^{(1)}$ is

$$i\frac{\partial\delta\Phi}{\partial t} + \frac{\partial^2\delta\Phi}{\partial x^2} + A\frac{\partial^2}{\partial x^2}[A(\delta\Phi + \delta\Phi^*)] + \left(A^2 - \kappa + \frac{d^2A^2}{dx^2}\right)\delta\Phi + A^2(\delta\Phi + \delta\Phi^*) = 0. \quad (7)$$

We seek localized solutions to this equation. By virtue of the symmetry of (1) under phase shift in the wave function, there exists a solution to (7) that is even in x and has the form $\delta\Phi = iA(x)$. Furthermore, the translational invariance of (1) implies that there exists an odd solution $\delta\Phi = dA/dx$. The existence of either "neutral" mode can readily be verified by substituting it into (7) and comparing the result with (3). The sought-for internal modes of a soliton are described by localized solutions to (7) oscillating in time:

$$\delta\Phi^{(1)}(x, t) = a[\psi(x)\exp(ipt) + \chi(x)\exp(-ipt)]. \quad (8)$$

Here, the eigenvalues p are supposed to be real and positive. The eigenfunctions $\psi(x)$ and $\chi(x)$ are determined by equations following from (7):

$$\begin{aligned} p_2\psi'' + p_1(\psi' + \chi') + (M - p)\psi + A^2\chi'' + q\chi &= 0, \\ p_2\chi'' + p_1(\chi' + \psi') + (M + p)\chi + A^2\psi'' + q\psi &= 0, \end{aligned} \quad (9)$$

where primes denote derivatives with respect to x,

$$\begin{aligned} p_1 = 2AA', \quad p_2 = 1 + A^2, \quad q = A^2 + AA'', \\ M = 2A^2 + 2(A')^2 + 3AA'' - \kappa. \end{aligned} \quad (10)$$

The boundary conditions for a discrete spectrum correspond to modes vanishing at infinity:

$$\psi(\pm\infty) = \chi(\pm\infty) = 0. \quad (11)$$

The eigenfunctions are normalized by the relation

$$\int_{-\infty}^{\infty} (\psi^2 + \chi^2)dx = 1, \quad (12)$$

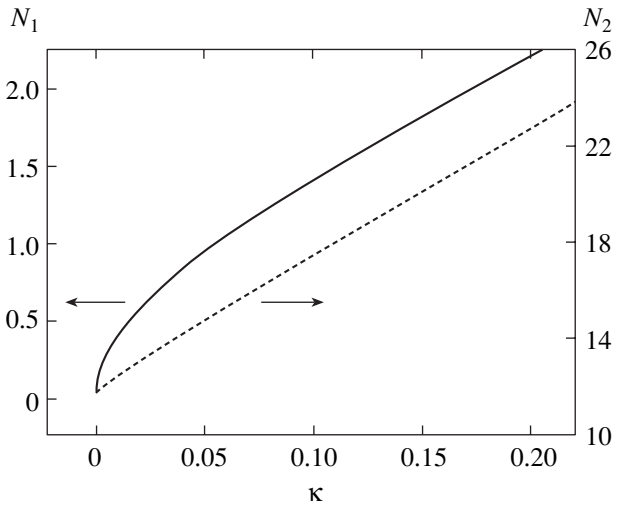


Fig. 2. Number of particles versus spectral parameter: N_1 and N_2 correspond to one- and two-dimensional solitons, respectively.

and a small value of the real amplitude a in (8) ensures that the perturbation is small.

Since the soliton is symmetric, we consider perturbations that are even and odd in x on the semi-infinite interval $0 < x < \infty$. It holds that

$$\psi'(0) = \chi'(0) = 0 \quad (13)$$

for even modes, while

$$\psi(0) = \chi(0) = 0 \quad (14)$$

for odd modes. The conditions at infinity are identical for modes of either parity:

$$\psi(\infty) = \xi(\infty) = 0. \quad (15)$$

More precisely, it follows from Eq. (9) that, if

$$0 < p < \kappa, \quad (16)$$

then both $\psi(x)$ and $\chi(x)$ exponentially decrease as $x \longrightarrow \infty$:

$$\begin{aligned} \psi(x) &\propto \exp(-\sqrt{\kappa + p}x), \\ \chi(x) &\propto \exp(-\sqrt{\kappa - p}x). \end{aligned} \quad (17)$$

According to (17), the function $\chi(x)$ decreases more slowly as compared to $\psi(x)$, particularly when $p \longrightarrow \kappa$, in which case the function $\psi(x)$ is localized within a finite interval, whereas the width of $\chi(x)$ increases indefinitely.

We found the internal modes for solitons with $0 < \kappa < 1000$ by using the tridiagonal algorithm to solve (9) subject to boundary conditions (15) combined with (13) or (14). The functions $\psi(x)$ and $\chi(x)$ corresponding to $\kappa = 300$ are shown in Figs. 1a and 1b for

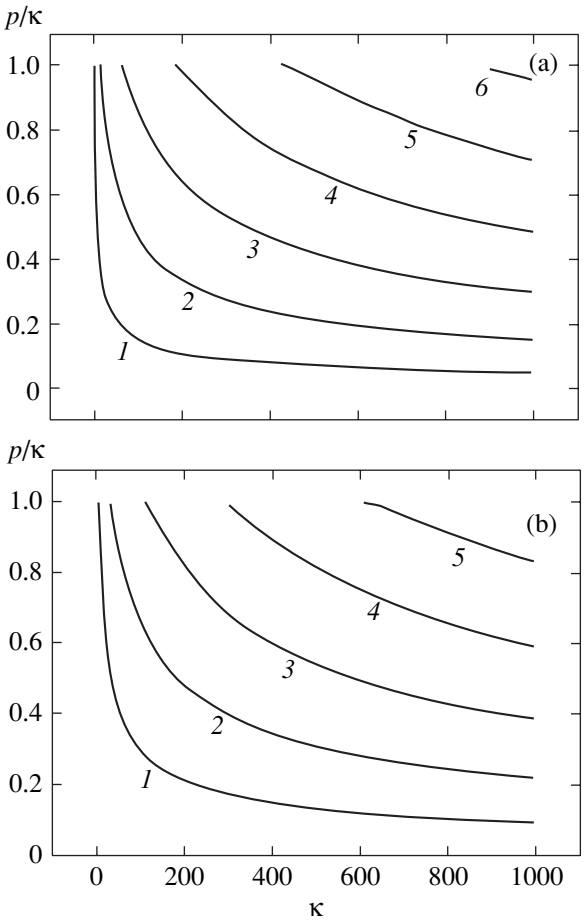


Fig. 3. Ratio of the internal-mode eigenfrequency p to the spectral parameter κ of a one-dimensional BEC soliton versus κ for (a) even and (b) odd modes. Numbers at curves are mode numbers.

two even modes and in Figs. 1c and 1d for two odd ones. The number of oscillation cycles in an eigenfunction increases with mode number. According to (17), the width of $\chi(x)$ increases as p approaches the continuous-spectrum boundary $p_b = \kappa$. The number of internal-mode branches increases with κ. Figure 3a shows that the first even mode exists when $\kappa > 0.4$; the second one, when $\kappa > 15$; the third, when $\kappa > 70$; and so on. The odd modes follow a similar pattern (see Fig. 3b). Note that neither even nor odd modes exist when $\kappa < 0.4$ and the interval of p for which they exist is bounded from above by $p_b = \kappa$.

3. NONLINEAR DAMPING OF INTERNAL MODES

In the linear approximation with respect to $\delta\Phi^{(1)}$, the amplitudes of oscillating perturbations localized in the neighborhood of a soliton do not decrease with time. The corresponding total wave functions describe undamped periodic oscillations of atom concentration in a BEC. However, when the terms nonlinear in $\delta\Phi^{(1)}$ are taken into account, the harmonics of the second and/or higher order (with frequencies shifted by $2p$ and so on) belong to the continuous spectrum corresponding to outward-propagating matter waves. This implies that internal modes are damped and a steady soliton is obtained in the long-time limit. In this section, the rate of nonlinear damping is found for the internal modes of one-dimensional BEC solitons by calculating the outward flow of atoms in a second-order perturbation theory. Since we use a procedure analogous to that developed in [13–17] for optical solitons, only a brief summary is presented here. Note also that the nonlinear damping mechanism does not work for collective oscillations of a BEC confined in a trap, because the trap prevents atoms from moving away to large distances from the condensate. The fact that BEC solitons have finite lifetimes (due to two- and three-body collisions resulting in spin flip or loss of condensate atoms) is ignored in this study, since they cannot be evaluated by solving the Gross–Pitaevskii equation. Some estimates concerning these processes were presented in [11, 12] for the (2,2) and (1,1) states of ^{7}Li atoms.

To introduce second-order terms in the perturbation amplitude, we write

$$\Phi(x, t) = [A(x) + \delta\Phi^{(1)}(x, t) + \delta\Phi^{(2)}(x, t)]\exp(i\kappa t), \tag{18}$$

where the superscript corresponds to the perturbation order. The second-order function can be sought in the form

$$\delta\Phi^{(2)}(x, t) = C(x)\exp(2ipt) + B^*(x)\exp(-2ipt) + C_0(x). \tag{19}$$

Generally, C, B, and C_0 are functions of time. In particular, they are responsible for nonlinear shift in the oscillation frequency p. However, the ensuing weak time dependence does not affect the internal-mode damping rate in question.

Substituting (19) into (1) and collecting the second-order terms, we obtain the following equations for $C(x)$ and $B(x)$:

$$\begin{aligned} &C'' - (2p + \kappa + 2A^2)C + A^2B + \psi(\psi + 2\chi)A \\ &\quad + A(BA)'' + C(A^2)'' + A(CA)'' \\ &\quad + \psi(A\psi)'' + \psi(A\chi)'' + A(\psi\chi)'' = 0, \\ &B'' + (2p - \kappa + 2A^2)B + A^2C + \chi(\chi + 2\psi)A \\ &\quad + A(CA)'' + B(A^2)'' + A(BA)'' \\ &\quad + \chi(A\psi)'' + \chi(A\chi)'' + A(\psi\chi)'' = 0. \end{aligned} \tag{20}$$

The function $C_0(x)$ describes both variation of perturbed soliton profile and nonlinear frequency shift. The corresponding evolution equation is not written out

here, because it is not coupled to the equations for $C(x)$ and $B(x)$ and does not contribute to the internal-mode damping rate. Representation (19) means that $C(x)$ and $B(x)$ are the amplitudes of time-periodic perturbations with period π/p. However, they exhibit substantially different behavior as functions of the coordinate. Indeed, as $|x| \longrightarrow \infty$, Eqs. (20) subject to the condition

$$2p - \kappa > 0 \tag{21}$$

yield an exponentially decaying $C(x)$ and a function $B(x)$ that tends to a plane-wave solution, $B(x) \sim \exp(\pm i\sqrt{2p - \kappa}\, x)$. Thus, $B(x)$ describes the damping of internal modes due to outward flow of atoms. According to [13–17], the time-dependent number of particles involved in an internal-mode perturbation,

$$\delta N = a^2 \int_{-\infty}^{\infty} (\psi^2 + \chi^2) dx = a^2,$$

is governed by the equation

$$\frac{d\delta N}{dt} = -\gamma \delta N^2, \tag{22}$$

where the nonlinear-damping rate constant is expressed as

$$\gamma = 2\pi\sqrt{2p - \kappa}\,|B(\infty)|^2. \tag{23}$$

The asymptotic value $B(\infty)$ is found by solving Eqs. (20) numerically.

Figures 4a and 4b show the numerical results obtained under condition (21) for even and odd modes, respectively. The highest values of the nonlinear-damping rate constant are several times greater for odd modes as compared to even ones. They are reached in the vicinity of $p_b = \kappa$. Since these values vary insignificantly between the internal-mode branches, several internal modes corresponding to the regions of overlapping curves in Fig. 4 may coexist over a relatively long time interval, whereas continuous-spectrum perturbations decay at higher rates. When relation (21) is violated, the internal modes are damped at much slower rates, because the outward flow of atoms is described by higher order perturbation theory.

4. INTERNAL MODES OF A TWO-DIMENSIONAL BOSE–EINSTEIN CONDENSATE

The two-dimensional modified Gross–Pitaevskii equation is obtained by replacing $\partial^2/\partial x^2$ in (1) with the transverse Laplace operator $\Delta_\perp = \partial^2/\partial x^2 + \partial^2/\partial y^2$. It was shown in [10–12] that stable two-dimensional bright BEC solitons with negative scattering lengths (independent of z) can exist even in the case of a weakly nonlocal interatomic interaction. Axially symmetric two-dimensional BEC solitons were obtained in [11]. Figure 2 illustrates the relation between the spectral parameter κ and the quantity

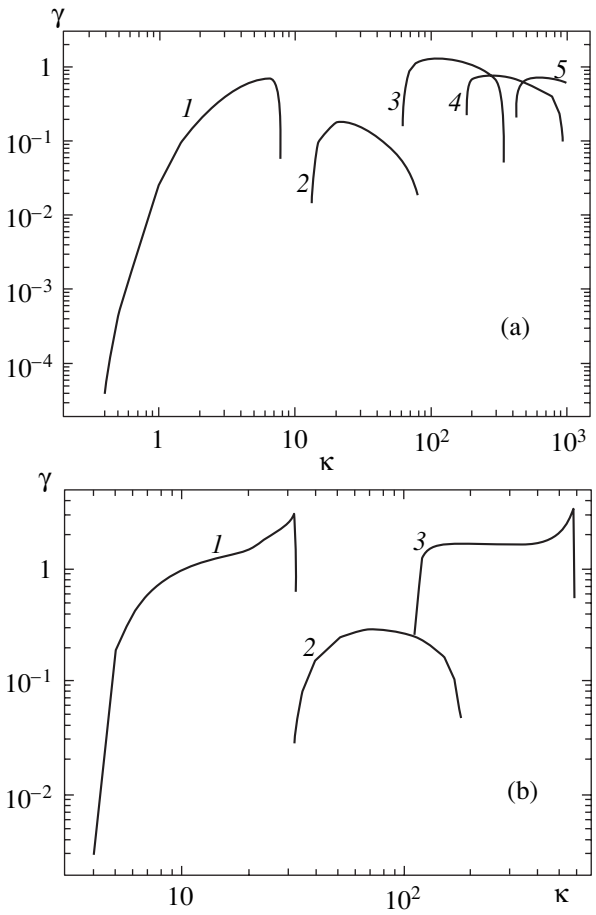


Fig. 4. Nonlinear damping rate constant versus the spectral parameter of a one-dimensional soliton for (a) even and (b) odd modes. Numbers at curves are mode numbers.

$$N = 2\pi \int_0^{\infty} A^2(r) r dr,$$

which is proportional to the number of particles. Figure 5a shows an example of radial profile of the soliton amplitude. In the general case, the internal modes are asymmetric and the following expansion of a particular mode in terms of angular harmonics can be used instead of (8):

$$\delta\Phi^{(1)}(r, \varphi, t) = a[\psi(r)\exp(im\varphi + ipt) + \chi^*(r)\exp(-im\varphi - ipt)], \tag{24}$$

where r and φ denote polar coordinates and the integer $m = 0, 1, 2, \ldots$ is an azimuthal index. The damping of internal modes of a two-dimensional soliton subject to

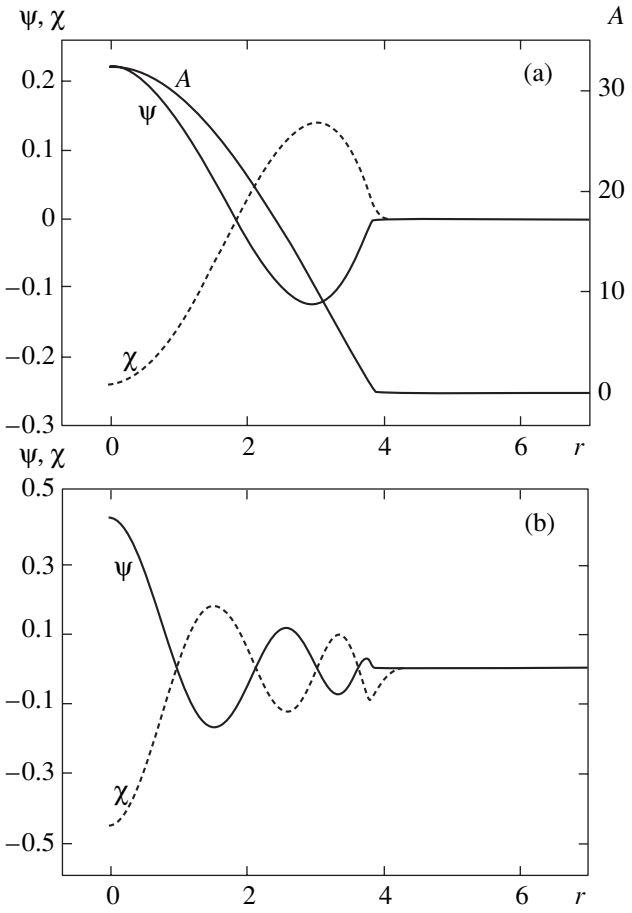


Fig. 5. Two-dimensional soliton with $\kappa = 300$: (a) amplitude $A(r)$ and eigenfunctions $\psi(r)$ and $\chi(r)$ of (a) the first and (b) the fourth symmetric internal modes.

condition (21) is also described by Eq. (22), with a damping rate constant γ determined by the asymptotic values of functions obtained in second-order perturbation theory [14–16]. Since the symmetric modes of two-dimensional solitons are the closest analogs of the even internal modes of one-dimensional solitons, the discussion that follows is restricted to the case of $m = 0$.

Omitting an analysis very similar to that presented above for one-dimensional solitons, we only illustrate the final results by Figs. 5–7. Comparing them with Figs. 1, 3, and 4, respectively, we find only a quantitative difference. The largest difference is in the damping rate constant: Figs. 4 and 7 demonstrate that it is substantially smaller in the two-dimensional case. The internal modes of spherically symmetric three-dimensional BEC solitons (which also exist when the interatomic interaction is nonlocal [11, 12]) have a more complicated angle-dependent structure than those described by (24). However, analogous results were obtained for spherically symmetric perturbations [14, 16].

Finally, recall that measurements of small-amplitude oscillation frequencies for trapped BECs are widely used to study the physics of interatomic interaction in condensates [1–6, 19]. The spectrum of collective excitations depends both on the properties of BEC atoms and on trap characteristics. Since the bright BEC soliton considered here is localized without any trap (along the "soliton coordinate"), we have a unique opportunity to examine the nonlocality of interatomic interaction in a condensate more directly by analyzing the eigenfrequencies of low-frequency internal modes. The strongest evidence of nonlocality was obtained for internal modes of the one-dimensional solitons implemented experimentally in [7, 8]. (Recall that no internal modes exist for the one-dimensional solitons described by the standard Gross–Pitaevskii equation; they exist only when the interatomic interaction is nonlocal.) Note that the soliton lifetime exceeds 3 s for lithium BECs [7], while the corresponding frequencies p lie in the kilohertz range. We believe that experimental stud-

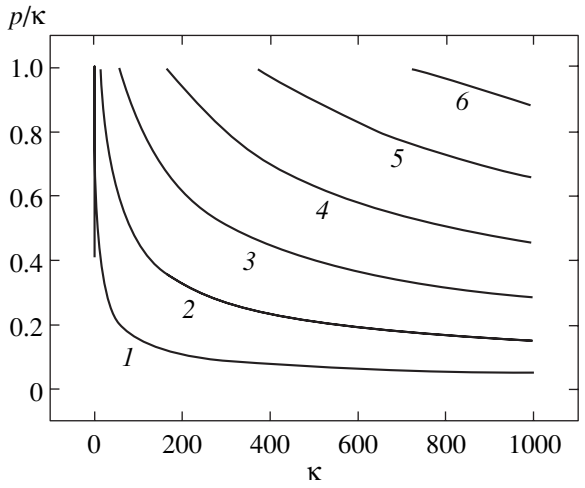


Fig. 6. Ratio of the eigenfrequency p for symmetric internal modes to the spectral parameter κ of a two-dimensional BEC soliton versus κ. Numbers at curves are mode numbers.

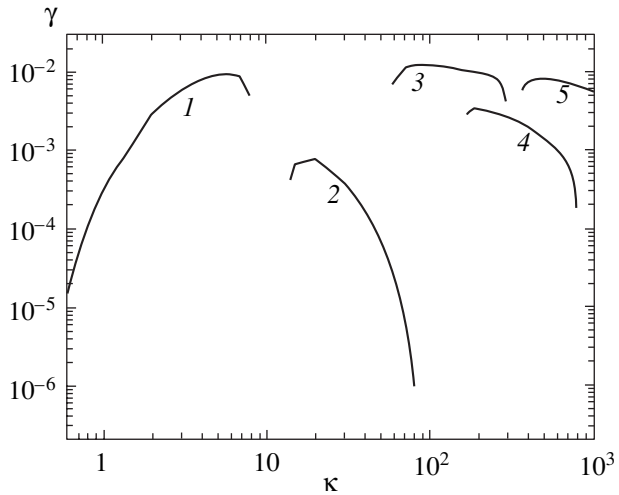


Fig. 7. Nonlinear damping rate constant versus the spectral parameter κ of a two-dimensional soliton. Numbers at curves are mode numbers.

ies of lithium BEC solitons can provide new information about nonlocal interaction between atoms in BECs. The internal-mode damping rate is determined by a weak outward flow of condensate atoms. This nonlinear mechanism does not work in the case of a trapped BEC. The damping rate strongly depends on the initial amplitude of atom-concentration perturbation, being very slow for small amplitudes. It is obvious that internal-mode damping can be ignored if the corresponding time scale exceeds the soliton lifetime determined by interatomic collisions.

ACKNOWLEDGMENTS

We thank V.A. Smirnov for helpful discussions. This work was supported by INTAS, project no. 211-855 and by the Russian Foundation for Basic Research, project no. 01-02-17242.

REFERENCES

1. L. P. Pitaevskiĭ, Usp. Fiz. Nauk **168**, 641 (1998) [Phys. Usp. **41**, 569 (1998)].
2. D. S. Jin, J. R. Ensher, M. R. Matthews, *et al.*, Phys. Rev. Lett. **77**, 420 (1996).
3. D. S. Jin, M. R. Mathews, J. R. Ensher, *et al.*, Phys. Rev. Lett. **78**, 764 (1997).
4. M.-O. Mewes, M. R. Andrews, N. J. van Druten, *et al.*, Phys. Rev. Lett. **77**, 988 (1996).
5. M. Edwards, P. A. Ruprecht, K. Burnett, *et al.*, Phys. Rev. Lett. **77**, 1671 (1996).
6. S. Stringari, Phys. Rev. Lett. **77**, 2360 (1996).
7. K. E. Strecker, G. B. Partridge, A. G. Truscott, and R. G. Hulet, Nature **417**, 150 (2002).
8. L. Khaykovich, F. Schreck, G. Ferrari, *et al.*, Science **296**, 1290 (2002).
9. S. K. Turitsyn, Teor. Mat. Fiz. **64**, 226 (1985).
10. N. N. Rosanov, A. G. Vladimirov, D. V. Skryabin, and W. J. Firth, Phys. Lett. A **293**, 45 (2002).
11. N. N. Rozanov, Yu. V. Rozhdestvenskiĭ, V. A. Smirnov, and S. V. Fedorov, Pis'ma Zh. Éksp. Teor. Fiz. **77**, 89 (2003) [JETP Lett. **77**, 84 (2003)].
12. N. N. Rozanov, Yu. V. Rozhdestvenskiĭ, V. A. Smirnov, and S. V. Fedorov, Opt. Spektrosk. **96**, 828 (2004) [Opt. Spectrosc. **96**, 757 (2004)].
13. D. E. Pelinovsky, Yu. S. Kivshar, and V. V. Afanasjev, Physica D (Amsterdam) **116**, 121 (1998).
14. N. N. Rosanov, S. V. Fedorov, N. A. Kaliteevskii, *et al.*, Nonlinear Opt. **23**, 221 (2000).
15. P. I. Krepostnov, V. O. Popov, and N. N. Rozanov, Opt. Spektrosk. **89**, 964 (2000) [Opt. Spectrosc. **89**, 887 (2000)].
16. A. B. Vitenberg, P. I. Krepostnov, V. O. Popov, and N. N. Rozanov, Opt. Spektrosk. **92**, 603 (2002) [Opt. Spectrosc. **92**, 550 (2002)].
17. N. N. Rosanov, P. I. Krepostnov, and V. O. Popov, Chaos **13**, 791 (2003).
18. N. N. Rozanov, Opt. Spektrosk. (2004) (in press).
19. W. Ketterle, Phys. Today **52** (12), 30 (1999).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 2, 2004, pp. 286–289.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 2, 2004, pp. 323–326.

ATOMS, SPECTRA, RADIATION

Resonance Internal Conversion in Hydrogen-Like Ions

F. F. Karpeshin[a], M. B. Trzhaskovskaya[b,*], and Yu. P. Gangrskii[c]

[a]*Institute of Physics, St. Petersburg State University, St. Petersburg, 198904 Russia*
[b]*St. Petersburg Institute of Nuclear Physics, Russian Academy of Sciences, Gatchina, 188300 Russia*
[c]*Joint Institute for Nuclear Research, Dubna, 141980 Russia*
**e-mail: trzhask@thd.pnpi.spb.ru*

Received June 2, 2003

Abstract—The resonance internal conversion effect in hydrogen-like ions is considered. For the $M1$ transition with an energy of 70.6 keV in ^{171}Yb, the lifetime of this nucleus in the excited state decreases by five orders of magnitude as a result of deexcitation via a new channel—the resonance conversion—provided that the nuclear transition energy is equal to the energy of transition of a single 1s electron to a higher ns state. Observation of the resonance conversion in hydrogen-like ions is a powerful method for investigating their nuclear structure.

1. INTRODUCTION

The resonance of a nucleus with an electron shell strongly influences the process of deexcitation of nuclear states [1–5]. In cases when the energies of electron and nuclear transitions are close, the energy can be transferred from nucleus to the electron shell. This phenomenon, called resonance, subbarrier, or discrete conversion, is rarely encountered in neutral atoms usually involved in experiments, which is explained by an extremely low probability of coincidence of the energies of atomic and nuclear levels. Indeed, vacant atomic levels in such atoms are concentrated in a very narrow (several electronvolts wide) energy interval near the ionization threshold and possess very small widths (~10^{-8} eV). However, the situation dramatically changes in ions with a single electron (hydrogen-like ions), where the interval of vacant levels expands up to several dozens of kiloelectronvolts and the level width may reach several electronvolts. In this situation, the probability of resonance conversion increases by many orders of magnitude.

The resonance conversion was experimentally discovered and theoretically studied in highly charged ions of ^{125}Te with a charge of $q \geq 45$ [6–8], where the subbarrier conversion takes place for 1s-shell electrons during the $M1$ nuclear transition with an energy of 35.492 keV. The 2p electrons produce damping of this resonance interaction, which has a twofold effect. On the one hand, this damping decreases the resonance conversion in magnitude. On the other hand, it was a rather strong damping that made the observation possible, since otherwise the probability of overlapping between nuclear and electron levels would be close to zero. The width of the level characterizing this damping was approximately 10 eV.

From this standpoint, the resonance conversion in hydrogen-like ions is of interest by providing a unique opportunity to observe this effect in the absence of damping. This ensures a very large (by many orders of magnitude) acceleration of deexcitation, since the effect is proportional to $(\Delta/\Gamma_a)^2$, where Δ is the resonance mismatch and Γ_a is the level width [1]. Indeed, for $\Delta \approx 1$ eV and $\Gamma_a \approx 10^{-3}$ eV, the gain reaches 5–6 orders of magnitude.

The resonance conversion can be observed in external magnetic or electric fields. A relatively large shift in the 1s level (on the order of several electronvolts) achieved for fields accessible in modern experiments, as well as the presence of a hyperfine structure, makes it possible to control the atomic transition frequency within an interval sufficiently broad for providing the exact resonance. Previously [1–3], we suggested using a laser tuned to the difference frequency for compensating the resonance mismatch.

Observation of the resonance conversion in hydrogen-like ions offers a powerful tool for investigating their nuclear structure. Using this technique, it is possible to determine the nuclear transition frequency to within the atomic level width (~10^{-3} eV) and to study the magnetic and electric moments of nuclei in both ground and excited states.

2. COMPUTATIONAL FORMULAS

The theory of resonance conversion has been developed in detail in [1–8]. The resonance conversion factor is defined by analogy with the conventional internal conversion coefficient, as the ratio of widths of the conversion and radiative transitions,

$$R = \frac{\Gamma_c}{\Gamma_\gamma}. \quad (1)$$

The R value can be calculated using the formula

$$R = \frac{\alpha_d \Gamma/2\pi}{\Delta^2 + (\Gamma/2)^2}, \quad (2)$$

where α_d is the resonance analog of the internal conversion coefficient. The latter quantity is calculated using expressions analogous to those for the conventional coefficient, but, in contrast to the latter, has the dimensionality of energy. The resonance frequency mismatch $\Delta = \omega_n - \omega_a$ is the difference between frequencies of the nuclear and electron transitions; Γ is the total width of the intermediate state, usually equal to the atomic level width.[1]

The standard situation is such that

$$\Delta/\Gamma \gg 1. \quad (3)$$

Relations (2) and (3) show that $R \ll 1$. However, as was noted above, the frequency of the electron transition can be controlled within certain limits so as to ensure exact resonance. In such cases (see below), $R \gg 1$. As suggested [1–3], this control can be provided by a laser. In the case under consideration, optical shift of the 1*s* electron level appearing in a constant external electric or magnetic field can be used. This method allows the transition energy to be controlled within several eV. In addition, it is necessary to take into account the hyperfine splitting [9].

3. LEVEL ENERGIES AND WIDTHS IN HYDROGEN-LIKE IONS

The energies of levels in hydrogen-like ions can be determined by solving the Dirac equation for an electron moving in the Coulomb field of the nucleus. The results of such calculations with allowance for the Lamb shift and finite size of the nucleus were reported in [10, 11]. Table 1 gives the energies of ns, $np_{3/2}$, $nd_{5/2}$, and $nf_{7/2}$ for a hydrogen-like ion of ^{171}Yb ($Z = 70$).

The existence of levels with angular momenta in a broad range in hydrogen-like ions admits the appearance of resonances for any multipolarity of nuclear transitions. As can be seen from Table 1, the range of transition energies featuring resonance is rather wide (approximately 18 keV for the ion of ^{171}Yb).

The widths of the levels under consideration are determined by the probabilities of $E1$ radiative transitions to lower states allowed by the selection rules. The lifetimes (or reduced widths) of levels in a hydrogen atom are presented in [12, 13]. In hydrogen-like ions, these widths depend on the atomic number (as Z^4) and on the principal quantum number (as n^{-3}). Examples of the level widths for the hydrogen-like ion of ^{171}Yb are presented in Table 2.

[1] We use the system of relativistic units in which $\hbar = c = m_e = 1$.

Table 1. The energies (keV) of electron levels in a hydrogen-like ion of ^{171}Yb [10, 11]

n	ns	$np_{3/2}$	$nd_{5/2}$	$nf_{7/2}$
1	0			
2	53.441	54.751		
3	63.738	64.125	64.237	
4	67.281	67.444	67.492	67.515
5			69.006	69.018
10		71.025	71.028	71.029

Table 2. The widths (eV) of electron levels in a hydrogen-like ion of ^{171}Yb [12, 13]

n	ns	$np_{3/2}$	$nd_{5/2}$	$nf_{7/2}$
2		9.88		
3	9.91×10^{-2}	2.94	1.02	
4	6.88×10^{-2}	1.28	0.434	0.217
5	4.39×10^{-2}	0.661	0.227	0.113
10	1.12×10^{-2}	0.0825	0.0282	0.0141
50	4.35×10^{-5}	0.00661	0.00227	0.00113

Table 3. Hyperfine splitting parameters for the ^{171}Yb nucleus

Nucleus	I^π	μ/μ_N	ΔW_F, eV	
			1*s*	10*s*
^{171}Yb	1/2	0.4397	0.569	5.69×10^{-4}

4. HYPERFINE SPLITTING OF ENERGY LEVELS

The hyperfine splitting of ns levels in hydrogen-like ions is determined primarily by the magnetic moment of the nucleus [14]:

$$\Delta W_F = \frac{4}{3}\alpha\left(\frac{\alpha Z}{n}\right)^3 \frac{\mu}{\mu_N}\frac{m_e}{m_p}\frac{2I+1}{2I}\left[1 + \frac{3}{2}(\alpha Z)^2\right]. \quad (4)$$

Here, μ is the magnetic moment of the nucleus, I is the nuclear spin, α is the constant of the fine structure, and m_e and m_p are the masses of electron and proton, respectively. An example of the hyperfine splitting parameters for the hydrogen-like ion of ^{171}Yb are presented in Table 3. These data indicate that the hyperfine

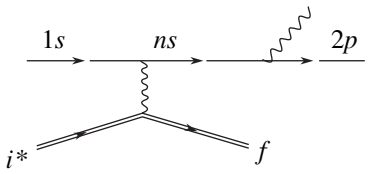


Fig. 1. The typical Feynman graph for the resonance internal conversion in a hydrogen-like ion. The double path corresponds to the nuclear transition from an excited state *i** to the ground state *f*.

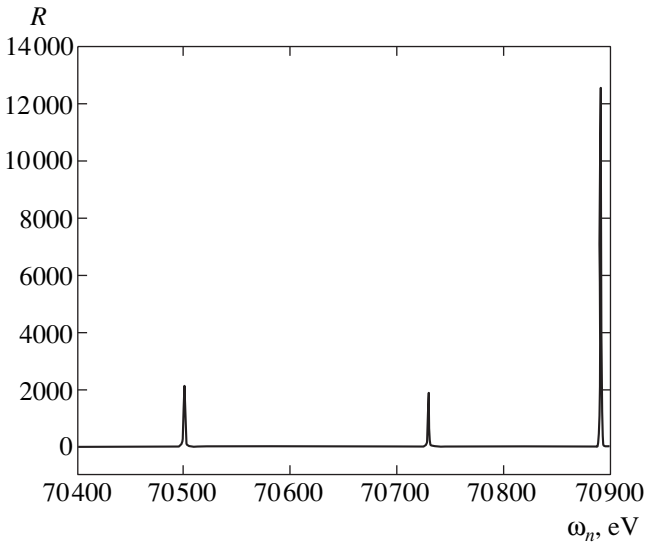


Fig. 2. Plot of the calculated *R* value versus nuclear transition energy ω_n in a hydrogen-like ion of ^{171}Yb in the vicinity of the *M*1 transition.

splitting significantly influences the position of the resonance.

The Zeeman splitting of electron levels is much smaller than the hyperfine splitting, since even the strongest magnetic fields (~10^4 Oe) are much weaker than internal atomic fields. For ^{171}Yb, the Zeeman splitting for $H = 10^4$ Oe is as small as 10^{-8} eV.

Table 4. The results of calculations of the electron transition energies ε, radiative transition widths Γ, and discrete internal conversion coefficients α_d for the *ns* electron levels of a hydrogen-like ion of ^{171}Yb in the vicinity of the *M*1 nuclear transition with $\omega_n \approx 70.6$ keV

ns	ε, keV	Γ, eV	α_d, keV
6*s*	69.646	0.0445	5.39
7*s*	70.166	0.0299	3.30
8*s*	70.501	0.0209	2.17
9*s*	70.729	0.0150	1.50
10*s*	70.892	0.0112	1.08

5. CALCULATIONS FOR THE *M*1 TRANSITION IN ^{171}Yb

The *M*1 nuclear transition energy in ^{171}Yb is 70.6 eV. The values of energies, radiative transition widths, and discrete internal conversion coefficients were calculated using the Dirac–Fock method with allowance for the finite size of the nucleus and for the higher QED corrections for the vacuum polarization, self-energy, etc. [11]. The computations were performed using the RAINE program package [15].

For the *M*1 transition, the main contribution to the resonance amplitude (Fig. 1) is due to the conversion electron transition to the *ns* state, for which the energy is close to that of the nuclear transition. In the resonance approximation, the contributions from other atomic states can be ignored taking into account relation (3) (see Table 4). In turn, the radiative transition widths are determined by the total probability of electric dipole transitions $ns \longrightarrow 2p, 3p, \ldots, (n-1)p$.

The results of calculations of the electron transition energies, radiative transition widths, and discrete internal conversion coefficients are presented in Table 4. Figure 2 shows a plot of the calculated *R* value versus nuclear transition energy in the 70.4–70.9 keV interval. As can be seen from these data, the coincidence of frequencies ($\omega_n = \omega_a$) provides for a discrete conversion factor on the order of 10^5.

6. CONCLUSIONS

The results of analysis of the resonance internal conversion for the *M*1 transition with an energy of 70.6 keV in ^{171}Yb showed that the resonance conversion factor *R* can reach several orders of magnitude (in the given case, 10^5), provided that the energies of the electron and nuclear transitions coincide. Under these conditions, the resonance conversion becomes the main deexcitation channel for the nucleus. The resonance can be observed by applying a constant external electric or magnetic field. The resonance mismatch can be also compensated using the corresponding resonance laser radiation field.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project nos. 99-02-17550 and 02-02-17117) and the U.S. Defense Threat Reduction Agency (contract no. DTRA01-01-P-0134).

REFERENCES

1. B. A. Zon and F. F. Karpeshin, Zh. Éksp. Teor. Fiz. **97**, 401 (1990) [Sov. Phys. JETP **70**, 224 (1990)].
2. F. F. Karpeshin, I. M. Band, M. B. Trzaskovskaya, and B. A. Zon, Phys. Lett. B **282**, 267 (1992).
3. F. F. Karpeshin, M. A. Listengarten, B. A. Zon, *et al.*, Can. J. Phys. **70**, 623 (1992).

4. F. F. Karpeshin, I. M. Band, and M. B. Trzhaskovskaya, Nucl. Phys. A **654**, 579 (1999).

5. F. F. Karpeshin, I. M. Band, M. B. Trzhaskovskaya, and M. A. Listengarten, Phys. Lett. B **372**, 1 (1996).

6. F. F. Karpeshin, M. R. Harston, F. Attallah, *et al.*, Phys. Rev. C **53**, 1640 (1996).

7. F. F. Karpeshin, I. M. Band, M. B. Trzhaskovskaya, *et al.*, Zh. Éksp. Teor. Fiz. **116**, 1565 (1999) [JETP **89**, 845 (1999)].

8. M. R. Harston, T. Carreyre, J. F. Chemin, *et al.*, Nucl. Phys. A **676**, 143 (2000).

9. F. F. Karpeshin, I. M. Band, M. B. Trzhaskovskaya, *et al.*, Phys. Rev. C **57**, 3085 (1998).

10. P. J. Mohr, At. Data Nucl. Data Tables **29**, 433 (1983).

11. W. R. Johnson and G. Soff, At. Data Nucl. Data Tables **33**, 406 (1985).

12. H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Academic, New York, 1957; Fizmatgiz, Moscow, 1960).

13. H. W. Kugel and D. E. Murnick, Rep. Prog. Phys. **40**, 297 (1977).

14. V. M. Shabaev, M. Tomaselli, T. Kuhl, *et al.*, Phys. Rev. A **56**, 252 (1997).

15. I. M. Band, M. A. Listengarten, M. B. Trzhaskovskaya, and V. I. Fomichev, Preprints Nos. 289 (1976), 298–300 (1977), 498 (1979), 1479 (1989) LIYaF AN SSSR (Inst. of Nuclear Physics, USSR Academy of Sciences, Leningrad).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 2, 2004, pp. 290–295.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 2, 2004, pp. 327–332.

ATOMS, SPECTRA, RADIATION

Relativistic Quantum Theory of Cyclotron Resonance in a Medium¶

H. K. Avetissian, G. F. Mkrtchian, and M. G. Poghosyan

Department of Quantum Electronics, Plasma Physics Laboratory, Yerevan State University, Yerevan, 375025 Armenia

e-mail: avetissian@ysu.am

Received October 20, 2003

Abstract—The relativistic quantum theory of cyclotron resonance in a medium with arbitrary dispersive properties is presented. The quantum equation of motion for a charged particle in the field of a plane electromagnetic wave and in the uniform magnetic field in a medium is solved in the eikonal approximation. The probabilities of induced multiphoton transitions between the Landau levels in a strong laser field are calculated. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

As is known, if a charged particle moves in the field of a transverse electromagnetic (EM) wave in the presence of a uniform magnetic field parallel to the direction of wave propagation, a resonant effect of the wave on the particle motion is possible. If the interaction takes place in vacuum, this is the well-known phenomenon of autoresonance [1–3], when the ratio of the Doppler-shifted wave frequency ω' to the cyclotron frequency Ω of the particle is conserved, $\omega'/\Omega = \text{const}$, and the resonance created at the initial moment automatically holds in the course of interaction. However, if the interaction takes place in a medium where the phase velocity of an EM wave is larger (a plasmalike medium) or smaller (a dielectric medium) than the speed of light in vacuum, the picture of the wave-particle interaction is substantially changed. In particular, the autoresonance phenomenon is violated in the medium because of a nonequidistant Stark shift of magnetic sublevels of an electron (Landau levels) in the electric field of an EM wave. As a result, the intensity effect of the wave governs the resonance characteristics and the particle state essentially depends on the initial conditions and the wave field magnitude at which the nonlinear resonance is achieved [4]. The cyclotron resonance (CR) in a medium was first investigated in the scope of the classical theory in papers [3, 5], where oscillating solutions for the particle energy were obtained. However, such behavior is valid only for an EM wave intensity less than a certain critical value. As shown in [4], at intensities above that critical value, a nonlinear resonance phenomenon of a threshold nature—so-called "electron hysteresis"—occurs (the EM wave is turned on adiabatically). If the intensity peak of an actual wave pulse exceeds the mentioned critical value, then significant acceleration of charged particles can be achieved (it is clear that the medium must be plasmalike for this purpose) [6].

Below the threshold intensity of electron hysteresis, when the linear CR occurs in a medium [3, 5], the free electron laser version has been proposed, based on the combined scheme of CR and Cherenkov radiation in a dielectric–gaseous medium [7].

We note that classical equations of motion for this process in a medium permit an exact solution only in a particular case where the initial velocity of a particle is parallel to the wave propagation direction and the wave has a circular polarization (namely, the electron hysteresis phenomenon has been obtained in this case).

Concerning the quantum description of CR, the relativistic quantum equation of motion allows exact solution only for CR in vacuum [8] (see [9] and references therein for a description of related quantum electrodynamic processes, such as electron–positron pair production, nonlinear Compton scattering in the presence of uniform magnetic field, etc., by this wave function). We note that the configuration of EM fields with a uniform magnetic field directed along the propagation of the transverse wave is one of the exotic cases in which the relativistic quantum equation of motion in vacuum allows an exact solution. In a medium, even in the absence of a uniform magnetic field, the relativistic quantum equation of motion for the particle–monochromatic wave interaction reduces to the Mathieu type (in general, Hill type) equation, the exact solution of which is unknown. In this case, obtaining an approximate analytic solution describing the nonlinear process of particle–wave interaction is already problematic [10–12].

The purpose of this paper is to obtain a nonlinear (in the field) approximate solution of the relativistic quantum equation of motion for a charged particle in the plane EM wave in a medium in the presence of a uni-

¶This article was submitted by the authors in English.

form magnetic field, a solution that describes sufficiently well the quantum picture of CR in a medium at high intensities of the external radiation field, in particular, multiphoton-stimulated transitions between Landau levels.

In what follows, the wave function of a charged particle moving in a medium in the field of a transverse EM wave in the presence of a uniform magnetic field directed along the wave propagation direction is obtained. Then the multiphoton CR in a medium is considered and the probabilities of induced multiphoton transitions in a strong circularly polarized EM wave are calculated.

2. WAVE FUNCTION OF A CHARGED PARTICLE IN THE PLANE ELECTROMAGNETIC WAVE IN A MEDIUM IN THE PRESENCE OF A UNIFORM MAGNETIC FIELD

Let a charged particle move in a medium in the field of a coherent EM wave and a uniform magnetic field along the wave propagation direction (z axis). The four-vector potential of this configuration of the EM field can be represented as

$$A_\mu(x) = A_\mu(x_1) + A_\mu(\tau), \tag{1}$$

where

$$A_\mu(x_1) = (0, x_1H_0, 0, 0) \tag{2}$$

is the four-vector potential of the uniform magnetic field with a strength $\mathbf{H}_0$ and

$$A_\mu(\tau) = \left\{A_1\left(t - n\frac{x_3}{c}\right), A_2\left(t - n\frac{x_3}{c}\right), 0, 0\right\} \tag{3}$$

is the four-vector potential of a plane transverse EM wave, x is the four-component radius vector, and

$$\tau = t - nx_3/c$$

is the plane wave coordinate. For four-component vectors, we chose the metric $a = (\mathbf{a}, ia_0)$. In (3), $n = n(\omega)$ is the refractive index of the medium and c is the speed of light in vacuum. Herein, we take the EM wave to be laser radiation that is quasimonochromatic with high accuracy ($\Delta\omega \ll \omega$, where ω is the carrier frequency),

$$n(\omega) \approx n = \text{const}.$$

We assume that the EM wave is switched on/off adiabatically, and therefore, for the vector potential $A_\mu(\tau)$, we have

$$A_\mu(\tau) = 0 \text{ at } t = \mp\infty.$$

Because we assume the coherent EM wave to be a laser radiation one for which the photon energy is negligibly small compared to the relativistic electron energy, we can neglect the spin interaction; the Dirac equation in quadratic form therefore reduces to the Klein–Gordon equation for a charged particle in (1),

$$\left\{\left(i\hbar\partial_\mu + \frac{e}{c}A_\mu(x)\right)^2 + m^2c^2\right\}\Psi(x) = 0, \tag{4}$$

where m and e are the particle mass and charge, respectively (we assume $e < 0$, with the electron in mind), and

$$\partial_\mu \equiv \frac{\partial}{\partial x_\mu}, \quad \mu = 1, 2, 3, 4,$$

denotes the first derivative of a function over the four-component radius vector x.

The particle quantum motion at $t \longrightarrow -\infty$, when $A_\mu(\tau) = 0$, is well known and has been the subject of numerous studies (see, e.g., [13]). In a uniform magnetic field, the particle motion is separated into the cyclotron (x_1, x_2) and the longitudinal (x_3) degrees of freedom. Because the coordinate x_2 is cyclic in this case (also in the presence of an EM wave; see (2) and (3)), the cyclotron motion is described by the set of quantum characteristics of the state $\{l, p_2\}$, where the number l denotes Landau levels ($l = 0, 1, 2, \ldots$) and p_2 is the x_2 component of the generalized momentum. The longitudinal motion at $t \longrightarrow -\infty$ is then described by the p_3 component of the particle initial momentum. Concerning the particle transverse initial state, we assume that at $t \longrightarrow -\infty$, the particle is in the $l = s$ Landau level. Therefore, the wave function of the particle at $t \longrightarrow -\infty$ is given by the known formula [13] (with the spin interaction neglected)

$$\Psi(x)|_{t\to-\infty} = N\exp\left[\frac{i}{\hbar}(p_3x_3 - E_s(p_3)t)\right]\Phi_{s,p_2}(x_\perp), \tag{5}$$

where N is the normalization constant, $x_\perp = \{x_1, x_2, 0, 0\}$, and

$$\Phi_{s,p_2}(x_\perp) = \exp\left(\frac{i}{\hbar}p_2x_2\right)U_s\left[\frac{x_1 + p_2a^2/\hbar}{a}\right], \quad a = \sqrt{\frac{\hbar c}{|e|H_0}}, \tag{6}$$

is the wave function corresponding to the cyclotron part of motion. Here, U_s are the Hermit functions, and the

dispersion law for the particle energy–momentum is

$$E_s^2(p_3) = m^2c^4 + c^2p_3^2 + 2|e|cH_0\hbar\left(s + \frac{1}{2}\right). \quad (7)$$

Because the EM wave field depends only on the retarded coordinate τ, it is more convenient to pass from the spacetime coordinates x_3, t to the wave coordinates

$$\tau = t - nx_3/c, \quad \eta = t + nx_3/c.$$

Then, due to the existence of a certain direction of the wave propagation, the variable η becomes cyclic, and hence the momentum conjugate to the coordinate η is conserved,

$$\frac{1}{2}\left(E_s(p_3) - \frac{c}{n}p_3\right) \equiv \Lambda = \text{const}. \quad (8)$$

This is the known integral of motion in this process according to the classical electrodynamics [4].

The particle wave function can then be sought in the form

$$\Psi(x) = \exp\left[-\frac{i}{\hbar}\Lambda\eta - \frac{i}{2\hbar}\left(E_s(p_3) + \frac{c}{n}p_3\right)\tau\right] f(x_\perp, \tau), \quad (9)$$

where the unknown function $f(x_\perp, \tau)$ is a slowly varying function of the variable τ compared with the exponential in (9). This approximation corresponds to the known eikonal approximation for the particle wave function, in which one can neglect the second derivative of $f(x_\perp, \tau)$ with respect to τ compared with the first-order derivative in equation of motion (4), which for the function $f(x_\perp, \tau)$ has the form

$$\left\{\frac{\hbar^2}{c^2}(n^2-1)\frac{\partial^2}{\partial\tau^2} + \frac{2i\hbar}{c^2}\tilde{E}\partial_\tau - \left(i\hbar\partial_\mu^\perp + \frac{e}{c}A_\mu(x)\right)^2 + \frac{E_s^2(p_3)}{c^2} - m^2c^2 - p_3^2\right\} f(x_\perp, \tau) = 0. \quad (10)$$

Here,

$$\partial_\mu^\perp = \{\partial_1, \partial_2, 0, 0\}, \quad \tilde{E} = E_s(p_3) - cnp_3.$$

We note that Eq. (10) is already a Hill-type equation even in the absence of a uniform magnetic field, and its exact solution is unknown. We therefore apply the eikonal approximation, considering $f(x_\perp, \tau)$ a slowly varying function of τ in Eq. (10) (the term with the second derivative of $f(x_\perp, \tau)$ describes the quantum recoil in the interaction of a particle with the EM wave), which is valid under the condition

$$\left|\frac{\hbar(n^2-1)}{2\tilde{E}}\frac{\partial^2 f}{\partial\tau^2}\right| \ll \left|\frac{\partial f}{\partial\tau}\right|. \quad (11)$$

Such an approximate solution describes the multiphoton interaction of particles with EM fields sufficiently well (for the electron–strong wave interaction in a medium, see [14]). Under condition (11), Eq. (10) implies the following equation for the function $f(x_\perp, \tau)$:

$$\left\{\frac{2i\hbar}{c^2}\tilde{E}\partial_\tau - \left(i\hbar\partial_\mu^\perp + \frac{e}{c}A_\mu(x)\right)^2 + \frac{E_s^2(p_3)}{c^2} - m^2c^2 - p_3^2\right\} f = 0. \quad (12)$$

In Eq. (12), the transverse and longitudinal motions are not separated. However, after a certain unitary transformation, the variables are separated [9]. The corresponding unitary transformation operator is

$$\hat{S} = \exp\{iK_\mu(\tau)\hat{P}_{\perp\mu}\},$$

$$\hat{P}_{\perp\mu} = -i\hbar\partial_\mu^\perp - \frac{e}{c}A_\mu(x_1), \quad (13)$$

$$K_\mu(\tau) = \{K_1(\tau), K_2(\tau), 0, 0\},$$

where $K_\mu(\tau)$ is chosen to separate the cyclotron and longitudinal motions and to satisfy initial condition (5), which is equivalent to the condition

$$K_1 + iK_2 = -\exp\left[-i\frac{ec}{\tilde{E}}H_0\tau\right] \times \int_{-\infty}^{\tau}\frac{ec}{\hbar\tilde{E}}(A_1(\tau') + iA_2(\tau'))\exp\left[i\frac{ec}{\tilde{E}}H_0\tau'\right]d\tau'. \quad (14)$$

For the transformed wave function $\tilde{f} = \hat{S}f(x_\perp, \tau)$, we then have the equation

$$\left\{\hat{P}_{\perp\mu}^2 - \frac{E_s^2(p_3)}{c^2} + p_3^2 + m^2c^2 - i\frac{2\hbar\tilde{E}}{c^2}\partial_\tau - \frac{e\hbar^2\tilde{E}}{c^3}F_{\nu\mu}K_\mu\frac{dK_\nu}{d\tau} + \left(\frac{e\hbar}{c}F_{\mu\nu}K_\nu - \frac{e}{c}A_\mu(\tau)\right)^2\right\}\tilde{f}(x_\perp, \tau) = 0, \quad (15)$$

where $F_{\mu\nu}$ is the EM field tensor corresponding to the uniform magnetic field $\mathbf{H}_0$. In Eq. (15), the variables are separated; by means of the inverse transformation $f = \hat{S}^+ \tilde{f}(x_\perp, \tau)$, we then obtain the solution of the initial equation (4) (taking Eq. (9) into account),

$$\Psi(x) = N\exp\left[\frac{i}{\hbar}(p_3x_3 - E_s(p_3)t) - \frac{i}{\hbar}\int_{-\infty}^{\tau} Q(\tau')d\tau'\right] \times \exp\left[i\frac{e}{c}H_0K_2\left(x_1 - \frac{\hbar}{2}K_1\right)\right]\Phi_{s,p_2}(x_1 - \hbar K_1, x_2 - \hbar K_2), \quad (16)$$

where

$$Q(\tau) = \frac{c^2}{2\tilde{E}}\left[\left(\frac{e\hbar}{c}F_{\mu\nu}K_\nu - \frac{e}{c}A_\mu(\tau)\right)^2 - \frac{e\tilde{E}\hbar^2}{c^3}F_{\nu\mu}K_\mu\frac{dK_\nu}{d\tau}\right]. \quad (17)$$

The obtained wave function (16) is valid under condition (11), which means that the particle total energy/momentum exchange occurring as a result of the multiphoton interaction with the strong EM wave at CR in a medium is much smaller than the initial energy/momentum of the particle. This energy/momentum exchange is determined by the full phase of wave function (16) with expressions (6), (14), and (17), which are found and estimated in the next section.

3. PROBABILITIES OF MULTIPHOTON TRANSITIONS BETWEEN LANDAU LEVELS

Although the particle motion in a uniform magnetic field is separated into cyclotron (x_1, x_2) and longitudinal (x_3) degrees of freedom, Eq. (5), these motions are not separated in the energy scale due to relativistic effects (7). For not very strong magnetic fields, however, we can separate the energies of longitudinal ($E_\parallel$) and cyclotron motions,

$$E_s(p_3) \approx E_\parallel + \hbar\Omega\left(s + \frac{1}{2}\right), \quad s\hbar\Omega \ll E_\parallel, \qquad \Omega = |e|cH_0/E_\parallel, \quad E_\parallel = \sqrt{m^2c^4 + c^2p_3^2}. \quad (18)$$

We now consider the concrete case of a circularly polarized quasimonochromatic EM wave with a main frequency ω and an average value $\bar{A}$ of the slowly varying envelope,

$$A_\mu(\tau) = \{-\bar{A}\sin(\omega\tau), g\bar{A}\cos(\omega\tau), 0, 0\}, \quad (19)$$

which is in resonance with the particle; i.e., the Doppler-shifted wave frequency is close to cyclotron,

$$\omega' \equiv (1 - nv_3/c)\omega \approx g\Omega, \quad (20)$$

where v_3 is the particle initial longitudinal velocity. In (19), the respective values $g = \pm 1$ correspond to the right- and left-hand circular polarizations of the wave. After the interaction ($t \longrightarrow +\infty$), under resonance condition (20), we have from Eq. (14)

$$K_1 + iK_2 = -\frac{e\bar{A}cT}{\hbar\tilde{E}}\exp(ig\omega\tau), \quad (21)$$

where T is the coherent interaction time (for a quasimonochromatic wave, $T \longrightarrow \infty$, and for actual laser radiation, T is the pulse duration).

The final state of the particle after the interaction is described by the wave function

$$\Psi_s(x) = N\exp\left[\frac{i}{\hbar}(p_3x_3 + p_2x_2 - E_s(p_3)t)\right] \times U_s\left[x_1 + \frac{e\bar{A}cT}{\tilde{E}}\cos(\omega\tau)\right] \times \exp\left[-i\frac{egH_0\hbar}{c}\left(\frac{e\bar{A}cT}{2\hbar\tilde{E}}\right)^2\sin(2\omega\tau) + i\frac{eg\bar{A}\Omega TE_\parallel}{\hbar c\tilde{E}}\left(x_1 + \frac{p_2a^2}{\hbar}\right)\sin(\omega\tau)\right]. \quad (22)$$

Expanding wave function (22) in terms of the complete basis of particle eigenstates (5),

$$\Psi_s(x) = \int dp_2'dp_3'\sum_{s'} C_{ss'}(p_2', p_3')\psi_{s', p_2', p_3'}(x), \quad (23)$$

we find the probabilities of multiphoton induced transitions between Landau levels.

To calculate the expansion coefficients $C_{ss'}(p_2', p_3')$, we use the result of the integration

$$\int_{-\infty}^{\infty} dz\exp(-ikz)U_s(a^{-1}z + ab)U_{s'}(a^{-1}z + ab') = \exp\{i\mu + i(s - s')\lambda\}I_{ss'}(\zeta), \quad (24)$$

where $I_{ss'}(\zeta)$ is the Lagger function and the characteristic parameters are determined by

$$\mu = \frac{ka^2(b + b')}{2}, \quad \lambda = \tan^{-1}\frac{k}{b' - b}, \qquad \zeta = a^2\frac{k^2 + (b - b')^2}{2}.$$

We then obtain the transition amplitudes

$$C_{ss'}(p_2', p_3') = \delta(p_2 - p_2')\delta(p_3 - p_3' - (s-s')g\omega n\hbar c^{-1}) \times \exp\left\{-\frac{i}{\hbar}(E_s(p_3) - E_{s'}(p_3') - (s-s')g\omega\hbar)t\right\} I_{ss'}[\zeta], \quad (25)$$

where $\delta(p)$ is the Dirac δ-function expressing the momentum conservation law and the argument of the Lagger function is

$$\zeta = \frac{e^2\bar{A}^2T^2\Omega E_\parallel}{2\hbar\tilde{E}^2}. \quad (26)$$

According to (25), the transition of a particle from an initial state $\{s, p_2, p_3\}$ to a state $\{s, p_2', p_3'\}$ is accompanied by emission or absorption of $s - s'$ photons. Consequently, substituting Eq. (25) into Eq. (23) and integrating over the momentum, we can rewrite the particle wave function in another form,

$$\Psi_s(x) = N\sum_{s'} I_{ss'}(\zeta) \times \exp\left\{-\frac{i}{\hbar}(E_s(p_3) - (s-s')g\omega\hbar)t + \frac{i}{\hbar}(p_3 - (s-s')g\omega n\hbar c^{-1})x_3 + \frac{i}{\hbar}p_2x_2\right\} U_{s'}(x_1). \quad (27)$$

The probability of the induced transition $s \longrightarrow s'$ between Landau levels is ultimately determined from formula (27):

$$w_{ss'} = I_{ss'}^2\left[\frac{e^2\bar{A}^2T^2\Omega E_\parallel}{2\hbar\tilde{E}^2}\right]. \quad (28)$$

Matching resonance condition (20) with formula (27) shows that in the field of a strong EM wave, the Landau levels are excited at the absorption of the wave quanta if $1 - nv_3/c > 0$ and $g = 1$, corresponding to the normal Doppler effect, while in the case where $1 - nv_3/c < 0$ and $g = -1$, which is possible in the refractive medium ($n > 1$), the Landau levels are excited at the emission of the wave quanta due to the anomalous Doppler effect.

We now estimate the average number of emitted (absorbed) photons by the electron at CR in a medium for highly excited Landau levels ($s \gg 1$). In accordance with the chosen approximation, the most probable number of photons in the strong EM wave field corresponds to the semiclassical limit ($|s - s'| \gg 1$), in which multiphoton processes dominate and the nature of the interaction process is very close to the classical one. In this case, the argument of the Lagger function can be represented as

$$\zeta = \frac{1}{4s}\left(\frac{\Delta E_{\rm cl}}{\hbar\omega}\right)^2, \quad \Delta E_{\rm cl} = \frac{e\mathscr{E}v_\perp T}{|1 - nv_3/c|}. \quad (29)$$

Here, $\Delta E_{\rm cl}$ is the amplitude of the energy change of the particle according to the classical perturbation theory, $\mathscr{E}$ is the amplitude of the electric field strength of the EM wave, and

$$v_\perp \approx c\sqrt{2\hbar s\Omega/E_{0\parallel}}$$

is the particle mean transverse velocity. The Lagger function is maximal at

$$\zeta \longrightarrow \zeta_0 = (\sqrt{s'} - \sqrt{s})^2,$$

exponentially falling beyond ζ_0. For the transition $s \longrightarrow s'$ with $|s - s'| \ll s$, we have

$$\zeta_0 \approx \frac{(s'-s)^2}{4s}.$$

Comparison of this expression with (28) and (29) shows that the most probable transitions are

$$|s - s'| \approx \frac{\Delta E_{\rm cl}}{\hbar\omega}, \quad (30)$$

in accordance with the correspondence principle. Using Eqs. (27) and (30), we can now represent the condition for the eikonal approximation in Eq. (11) as

$$\Delta E_{\rm cl} \ll 2\left|\frac{E_s(p_3) - cnp_3}{n^2 - 1}\right|.$$

This condition actually restricts the intensity of the EM wave field in accordance with Eq. (29). However, the above condition is practically very weak, and the wave function obtained in (16) describes multiphoton transitions at the CR in strong laser fields with great accuracy.

4. CONCLUSIONS

In the scope of relativistic quantum theory, a nonlinear (in the field) wave function of the eikonal type of a charged particle in the plane EM wave and a uniform magnetic field in a medium has been obtained neglecting spin interaction and, consequently, quantum recoil of photons (in accordance with the eikonal approximation applied). The eikonal approximation does not restrict hardly at all the applicability of such a wave function in the actual cases of strong radiation fields that are laser fields (with photon energy much smaller than electron energy). This wave function describes

sufficiently well the quantum picture of CR in a medium at high intensities of the external radiation field, in particular, multiphoton stimulated transitions between Landau levels.

With this wave function, one can treat a large class of nonlinear quantum electrodynamic processes in strong EM fields with the modifications that a medium offers (e.g., the anomalous Doppler effect), including astrophysical applications, where CR plays a significant role [15]. In addition, one of the advantages of CR in a dielectric medium is that for a moderate relativistic particle beam, one can achieve the CR in the optical region (close to the Cherenkov resonance) by current lasers and existing uniform magnetic fields (~10^4 Gs), while in vacuum, CR with radio frequencies is possible with the same parameters. Finally, the wave function obtained is especially important for describing the radiation process by a charged particle at CR in gaseous media consisting of a superposition of Compton, Cherenkov, and synchrotron radiation.

We note that radiation of a particle undergoing laser-assisted multiphoton transitions at CR between Landau levels has already been investigated and will be presented elsewhere.

ACKNOWLEDGMENTS

This work was supported by the International Science and Technology Center (ISTC) (project no. A-353) and by the National Foundation of Science and Advanced Technologies (grant no. NFSAT PH-082-02/CRDF-12023).

REFERENCES

1. A. A. Kolomensky and A. N. Lebedev, Dokl. Akad. Nauk SSSR **145**, 1259 (1962) [Sov. Phys. Dokl. **7**, 745 (1963)].
2. V. Ya. Davidovsky, Zh. Éksp. Teor. Fiz. **43**, 886 (1962) [Sov. Phys. JETP **16**, 629 (1963)].
3. A. A. Kolomensky and A. N. Lebedev, Zh. Éksp. Teor. Fiz. **44**, 261 (1963) [Sov. Phys. JETP **17**, 179 (1963)].
4. V. M. Haroutunian and H. K. Avetissian, Izv. Akad. Nauk Arm. SSR, Fiz. **9**, 110 (1975).
5. C. S. Roberts and S. J. Buchsbaum, Phys. Rev. A **135**, 381 (1964).
6. H. K. Avetissian, in *Abstracts of the ICOMP V* (Paris, 1990).
7. H. K. Avetissian and K. Z. Hatsagortsian, Zh. Tekh. Fiz. **54**, 2347 (1984) [Sov. Phys. Tech. Phys. **29**, 1381 (1984)].
8. P. J. Redmond, J. Math. Phys. **6**, 1163 (1965).
9. I. M. Ternov, V. R. Khalilov, and V. N. Rodionov, *Interaction of Charged Particles with Strong Electromagnetic Field* (Mosk. Gos. Univ., Moscow, 1982) [in Russian].
10. H. K. Avetissian *et al.*, Zh. Éksp. Teor. Fiz. **113**, 43 (1998) [JETP **86**, 24 (1998)].
11. H. K. Avetissian *et al.*, Phys. Lett. A **244**, 25 (1998).
12. H. K. Avetissian *et al.*, Phys. Lett. A **246**, 16 (1998).
13. A. I. Akhiezer and V. B. Berestetskiĭ, *Quantum Electrodynamics*, 3rd ed. (Nauka, Moscow, 1969; Wiley, New York, 1965).
14. H. K. Avetissian, Phys. Lett. A **58**, 144 (1976).
15. V. L. Ginzburg, *Theoretical Physics and Astrophysics* (Nauka, Moscow, 1957; Pergamon Press, Oxford, 1979).

Journal of Experimental and Theoretical Physics, Vol. 99, No. 2, 2004, pp. 296–310.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 2, 2004, pp. 333–348.

ATOMS, SPECTRA, RADIATION

A New Approach to the Theory of Cherenkov Radiation Based on Relativistic Generalization of the Landau Criterion

S. G. Chefranov

Obukhov Institute of Atmospheric Physics, Russian Academy of Sciences, Moscow, 109017 Russia
Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences, Pushchino, Moscow oblast, 142290 Russia
e-mail: schefranov@mail.ru

Received November 14, 2002; in final form, February 20, 2004

Abstract—Relativistic generalization of the Landau criterion is obtained which, in contrast to the classical Tamm–Frank and Ginzburg theories, determines the primary energy mechanism of emission of nonbremsstrahlung Cherenkov radiation. It is shown that Cherenkov radiation may correspond to a threshold energetically favorable conversion of the condensate (ultimately long-wavelength) elementary Bose perturbations of a medium into transverse Cherenkov photons emitted by the medium proper during its interaction with a sufficiently fast charged particle. The threshold conditions of emission are determined for a medium with an arbitrary refractive index n, including the case of isotropic plasma with $n < 1$ for which the classical theory of Cherenkov radiation prohibits such direct and effective nonbremsstrahlung emission of these particular transverse high-frequency electromagnetic waves. It is established that these conditions of emission agree with the data of well-known experiments on the threshold for observation of Cherenkov radiation, whereas the classical theory only corresponds to the conditions of observation of the interference maximum of this radiation. The possibility of direct effective emission of nonbremsstrahlung Cherenkov radiation, not taken into account in the classical theory, is considered for many observed astrophysical phenomena (type III solar radio bursts, particle acceleration by radiation, etc.).

1. INTRODUCTION

The possibility that a non bremsstrahlung mechanism of emission can be operative in nature during uniform motion of a sufficiently fast charged particle in a refractive medium had been theoretically predicted by Heaviside [1, 2] half a century before Cherenkov radiation was experimentally discovered [3, 4] and a macroscopic theory of this radiation was created (independently of [1]) by Tamm and Frank [5–9]. The phenomenological quantum theory of Cherenkov radiation proposed by Ginzburg [10, 11] did not actually modify the main conclusions of the Tamm–Frank theory. According to the classical theory [5–11], anisotropic Cherenkov radiation is emitted directly by an electron moving in a medium at a velocity v_0 greater than the phase velocity of light waves in this medium, that is, at $v_0 > c/n$ (where c is the speed of light in vacuum and n is the refractive index of this medium) and only for $n > 1$. However, the concept of photon emission directly from the moving electron disagrees with the essence of the mechanism of Cherenkov radiation, as indicated by Tamm in [6] (see also [12]): "From the standpoint of microscopic theory, this radiation is not emitted directly by an electron, but arises due to the coherent oscillations of molecules of the medium excited by this electron. However, we will not dwell here on a microscopic consideration of the problem." This conceptual discrepancy, considered in more detail below (see Sections 1.1 and 1.2), is responsible for the fact that there is still no clear understanding of the principles of the energy mechanism of Cherenkov photon emission by a refractive medium interacting with a fast charged particle.

1.1. Proceeding from Maxwell's equations for a refractive medium with $n > 1$, Tamm [6] showed that the field generated by an electron moving uniformly in the medium at a velocity of v_0 ceases to decay exponentially with distance from the axis of the particle motion only when $v_0 > c/n$. This result is not immediately indicative of the possible appearance of a cone-shaped anisotropic distribution of the electron-induced field, which is characteristic of Cherenkov radiation. In making a conclusion about this possibility, we implicitly use [6] an additional a priori assumption that a coherent radiation exists in the medium for $v_0 > c/n$. Under this assumption, the interference theory leads to anisotropic distribution in the interference maximum of the field determined by the relation $\cos\theta_0 = c/nv_0$ (see [6, Eq. (2.1)]), where θ_0 is the angle between vector $\mathbf{v}_0$ and the direction from the electron to the interference maximum in the distribution of Cherenkov radiation intensity. Without recourse to interference theory, an aniso-

tropic cone-shaped distribution of the electron-induced field (for the same θ_0 and $v_0 > c/n$) coinciding with the Heaviside solution [1, 2] was obtained in [6] only in the case of n = const (i.e., in the absence of dispersion determining the dependence of n on the wave frequency ω). However, experiments [3, Ref. 2] showed that the limiting opening angle θ_m of the anisotropic cone of Cherenkov radiation corresponds to much greater angles θ_0. It is the θ_m (rather than θ_0) value that determines a threshold for the Cherenkov radiation emission, which is therefore not established in the classical theory of this radiation [5–9]. In addition, it is known [13] that neglect of the frequency dispersion of medium is absolutely unacceptable near the threshold of Cherenkov photon emission.

Therefore, the condition of Cherenkov radiation emission formulated as $v_0 > c/n$ for $n > 1$ in the Tamm–Frank theory [5–9] does not in fact describe the necessary threshold conditions. These conditions have to be related to changes in the energy state of the medium (not taken into account in [5–9]) for which the emission of Cherenkov photons by the medium itself during interaction with a sufficiently fast electron.

1.2. The quantum theory of Cherenkov radiation proposed by Ginzburg [10, 11] also does not take into account the aforementioned changes in the energy state of the medium necessary for the emission of Cherenkov photons. In the Ginzburg approach, in contrast to the Tamm–Frank theory, the conclusion about the anisotropy of Cherenkov radiation is obtained directly within the framework of the phenomenological quantum theory of this radiation [10, 11], without implicitly using (as in [6]) additional considerations of the interference theory. Moreover, the Ginzburg theory [10, 11] even takes into account, although implicitly, a certain kind of medium response to Cherenkov radiation. However, this response is reflected only in the momentum balance equation [10, 11] and is not introduced into the energy balance equation.

In the Ginzburg theory [10, 11], as well as in the Tamm–Frank theory [5–9], Cherenkov photons are actually emitted directly by a fast moving electron. However, within the framework of relativistic theory, it is impossible to simultaneously meet the laws of conservation of energy and momentum in description of the process of photon emission (or absorption) directly by a free electron moving in vacuum [14, 15] where a real photon has zero rest mass. In the Ginzburg theory [10–11], this restriction is removed by taking into account the presence of a medium, which plays the role of a third body accepting the excess momentum and/or a factor providing for the appearance of a nonzero effective rest mass of the Cherenkov photon (due to refractive properties of the medium). According to this theory, a photon in the medium with $n > 1$ is characterized by the Minkowski pseudomomentum

$$\mathbf{p}_m = \frac{\varepsilon_p n}{c}\mathbf{k},$$

where ε_p and v_p are the energy and velocity of the Cherenkov photon, respectively, and $\mathbf{k} = \mathbf{v}_p/v_p$ is a unit vector whose direction in the isotropic medium coincides with that of the wave vector of Cherenkov radiation. Assuming that $\mathbf{p}_m$ corresponds to the intrinsic momentum $\mathbf{p}$ of a photon in the medium and using the well-known relativistic relation between ε_p, $\mathbf{p}$, and the rest mass m_p [16],

$$m_p^2c^2 = \frac{\varepsilon_p^2}{c^2} - \mathbf{p}^2,$$

we can formally determine for $\mathbf{p} = \mathbf{p}_m$ a nonzero (but complex) rest mass of the Cherenkov photon as $m_p = \tilde{m}_p$, where

$$\tilde{m}_p^2 = -\frac{\varepsilon_p^2}{c^4}(n^2 - 1) < 0.$$

The quantity $\tilde{m}_p^2 < 0$ was considered, for example, in the description of Cherenkov radiation in a strong electromagnetic field [17]. Such a field plays the role of a refractive medium [15] with $n \longrightarrow 1$ at $n > 1$. Although the complex mass was not explicitly introduced in [10, 11], all conclusions of the phenomenological quantum Ginzburg theory are valid only under the condition that $\tilde{m}_p^2 \neq 0$. Indeed, taking into account the above representation of $\tilde{m}_p^2 < 0$, formula [11, Eq. (11)] describing the condition of Cherenkov photon emission in a medium with $n > 1$can be rewritten exactly as

$$1 \geq \cos\theta = \frac{c}{v_0\sqrt{1 - \tilde{m}_p^2c^4/\varepsilon_p^2}}\left(1 - \frac{\tilde{m}_p^2c^2}{2m_e\varepsilon_p\Gamma_0}\right), \quad (1)$$

where $\Gamma_0 = (1 - v_0^2/c^2)^{-1/2}$, m_e is the electron mass, and θ is the angle between vectors $\mathbf{v}_0$ and $\mathbf{k}$. Obviously, inequality (1) is violated in the limit when $\tilde{m}_p^2 \longrightarrow 0$ at fixed ε_p and v_0.

As is known [10], the pseudomomentum $\mathbf{p}_m$ is not equal to the intrinsic momentum of a photon in a medium, but can only be related to it as $\mathbf{p}_m = \mathbf{p}_a + \mathbf{p}_c$ (see also the discussion concerning this relation [18, 19]), where $\mathbf{p}_a$ is the true intrinsic photon momentum given by the well-known Abragam formula,

$$\mathbf{p} = \frac{\varepsilon_p\mathbf{v}_p}{c^2} = \mathbf{p}_a$$

($\mathbf{p}_a = \varepsilon_p\mathbf{k}/nc$ for $v_p = c/n$ at $n > 1$; $\mathbf{p}_a = \varepsilon_p n\mathbf{k}/c$ for $v_p = cn$ at $n < 1$), and $\mathbf{p}_c$ is the Abragam force impulse trans-

ferred by photon to the medium. For example, in a medium with $n > 1$ [10]

$$\mathbf{p}_c = \frac{\varepsilon_p(n^2-1)}{nc}\mathbf{k}.$$

For $\mathbf{p} = \mathbf{p}_a$, the rest mass of a Cherenkov photon is not only nonzero, but always a real quantity for any $n \neq 1$ since

$$m_p^2 = \frac{\varepsilon_p^2(n^2-1)}{c^4 n^2}, \quad n > 1,$$

$$m_p^2 = \frac{\varepsilon_p^2}{c^4}(1-n^2), \quad n < 1.$$

However, this circumstance was not taken into account in [10, 11] and no correction was introduced into the energy balance equation for Cherenkov radiation in order to take into account the change in the energy state of the medium (whose internal energy decreases, at least by $m_p c^2$, upon emission of the Cherenkov photon).

Thus, the classical theory [5–11], considering Cherenkov radiation as emitted directly by a moving electron, does not take into account the necessary change in the energy state of the medium emitting Cherenkov photons with $m_p^2 > 0$. Therefore, this theory cannot determine the threshold conditions necessary for the emission of Cherenkov photons and leaves questions concerning the primary energy mechanism and the physical nature of nonbremsstrahlung Cherenkov radiation unanswered.

1.3. A theory proposed in this paper develops, in contrast to the classical Tamm–Frank and Ginzburg theories [5–11], the aforementioned concept, according to which Cherenkov photons with $m_p^2 > 0$ are emitted by the medium itself, while a sufficiently fast charged particle only initiates this emission by interacting with the medium. Using this approach, the necessary threshold conditions for the emission of Cherenkov radiation are established based on the obtained relativistic generalization of the Landau criterion [20]. It will be demonstrated that the phenomenon of nonbremsstrahlung Cherenkov radiation offers an example of the well-known mechanism of dissipative (secular) instability [21–24], whereby the emission of a "massive" Cherenkov photon by a medium, accompanying the absorption of a condensate (ultimately long-wavelength) elementary Bose perturbation in the medium, becomes energetically favorable at over-threshold electron velocities $v_0 > c/n_*(n)$, where $n_* > 1$ and $n_* > n$ for $n > 1$ and $n_* > 1$ for $n < 1$. A change in the energy state of the medium during Cherenkov radiation is determined by the absorbed energy quantum $\varepsilon_0 = \hbar\tilde{\omega}_0$ of the Bose excitation, which can be of thermal (for $\tilde{\omega}_0 \sim kT/\hbar$) or other physical nature (in plasma, $\tilde{\omega}_0$ is equal to the plasma frequency ω_0).

If the energy ε_p of the emitted Cherenkov photon is greater than the energy of absorbed excitation quantum, then (e.g., for high-frequency Cherenkov radiation in the plasma with $n < 1$) the energy of the absorbed long-wavelength plasmon ($\tilde{\mathbf{k}} \longrightarrow 0$, $\mathbf{p}_0 = \hbar\tilde{\mathbf{k}}$ is the momentum of the Bose excitation) may precisely determine the aforementioned rest mass m_p of the Cherenkov photon since $m_p = \hbar\omega_0/c^2$. Indeed, this relation and the above definition of m_p for $n < 1$ gives the well-known law of dispersion for transverse plasmons,

$$n^2 = 1 - \frac{\omega_0^2}{\omega^2}.$$

For these plasmons, the dispersion equation $\omega^2 = \omega_0^2 + k^2c^2$ coincides with the above relativistic relation between ε_p, $\mathbf{p}$, and m_p for $\varepsilon_p = \hbar\omega$ and $\mathbf{p} = \mathbf{p}_a = \hbar\mathbf{k}$. In the general case, $m_p c^2$ may differ from ε_0. For example, as will be shown below for $\varepsilon_0 > \varepsilon_p$, it is possible that the emission of a Cherenkov photon is even accompanied by an additional increase in the kinetic energy of an electron. Thus, as was already noted above and pointed out by Zrelov [13], the threshold for the Cherenkov radiation emission significantly depends on the dispersion properties of the refractive medium. The necessary conditions established in this study for the emission of Cherenkov photons show a better qualitative and quantitative agreement with available experimental data [3] on the limiting opening angles θ_m of the Cherenkov radiation cone as compared to the predictions of the classical Tamm–Frank and Ginzburg theories.

Below, we interpret the known data obtained by radio astronomy observations and the physics of cosmic rays in terms of the proposed theory of nonbremsstrahlung Cherenkov radiation. We also compare predictions of the new theory with the conclusions of some other theoretical investigations in the direction under consideration [25–27], which were inconsistent in using only the vacuum representation of the photon momentum (i.e., $|\mathbf{p}| = \varepsilon_p/c$ at $m_p = 0$) in studying (apart from the context of nonbremsstrahlung Cherenkov radiation) phenomena involving the emission of high-energy quanta under the conditions when fast particles absorbed elementary excitations of a refractive medium with $n \neq 1$. In particular, it was concluded [26] that such emission might represent a kind of bremsstrahlung radiation emitted as a result of conversion of relatively short-wavelength longitudinal plasmons (with a phase velocity lower than c) into transverse plasmons during the Compton effect on plasma waves for relativistic electrons. However, no particular values of electron velocities were indicated in [25, 26] as emission thresholds, and the anisotropy of such radiation was even not mentioned.

As will be shown below, the aforementioned conclusion [26] concerning the mechanism of emission is not valid in the case of absorption of ultimately long-wavelength (condensate) elementary Bose perturbations (including longitudinal plasmons) considered in this study (and in [25]), which corresponds exactly to a threshold (with respect to the electron velocity) emission of anisotropic nonbremsstrahlung Cherenkov radiation. Such radiation may also correspond to the Compton effect of a graviton on a fermion [28], whereby this condensate graviton (representing a long-wavelength elementary Bose perturbation of a medium) is converted into a photon and the gravitational field serves as the medium.

2. QUANTUM THEORY OF CHERENKOV RADIATION

The physically consistent quantum theory of nonbremsstrahlung Cherenkov radiation formulated below establishes, in contrast to the classical theory [5–11], threshold changes in the state of a refractive medium interacting with a uniformly moving charged particle, which are necessary for the emission of such radiation. Together with the relativistic generalization of the Landau criterion [20] (obtained in Section 3), this theory elucidates the primary physical mechanism of emission of nonbremsstrahlung Cherenkov radiation by the refractive medium. The consideration below is performed using the terms and notation as given in Section 1.

2.1. Consider a system comprising a medium of mass M and a charged particle (e.g., electron) of mass m_e. In the initial state, the medium is at rest while the particle performs rectilinear uniform motion in the laboratory frame at a velocity of $\mathbf{v}_0$. By analogy with [20], it is assumed that dissipative factors (determined, e.g., by a polarization interaction between the moving electron and the medium) make possible the energetically favorable threshold creation of an elementary excitation in the medium, which has the form of a Cherenkov photon propagating at a velocity of $\mathbf{v}_p$. Let the photon have energy ε_p and momentum

$$\mathbf{p} = \frac{\varepsilon_p \mathbf{v}_p}{c^2} = \mathbf{p}_a,$$

where $v_p \equiv |\mathbf{v}_p| \le c$. The electron, even moving outside a medium (for Cherenkov radiation in this case, see [11]) can play the same role as that of the walls of a capillary with liquid helium, initiating the above-threshold (with respect to the velocity of liquid relative to the walls) creation of the vortex ("roton") perturbations breaking superfluidity in the helium flow as a result of viscous interaction between the solid boundary and the flow [20].

In order to determine the threshold velocity $\mathbf{v}_0$ at which the emission of a Cherenkov photon becomes energetically favorable, it is possible to restrict the consideration (as in [20]) to the energy balance equation (see Section 3). However, for convenience of comparison with the Ginzburg quantum theory [10, 11] and with the theory of inverse Compton effect [25, 26], we will retain the complete system, including the energy and momentum equations describing Cherenkov radiation with $\varepsilon_p > 0$ with allowance for possible changes in the state of both the electron and the medium in the laboratory reference frame:

$$\begin{aligned} &m_e c^2 \Gamma_0 + Mc^2 \\ &= (M - \Delta M)c^2 \Gamma_2 + \varepsilon_p + m_e c^2 \Gamma_1 + \Delta U, \end{aligned} \tag{2}$$

$$m_e \mathbf{v}_0 \Gamma_0 = (M - \Delta M)\mathbf{v}_2 \Gamma_2 + \frac{\varepsilon_p \mathbf{v}_p}{c^2} + m_e \mathbf{v}_1 \Gamma_1. \tag{3}$$

Here, $\Gamma_\alpha = (1 - v_\alpha^2/c^2)^{-1/2}$ ($\alpha = 0, 1, 2$) and ΔMc^2 is a change in the internal energy of the medium as a result of emission of a Cherenkov photon with energy $\varepsilon_p = \hbar\omega$ and momentum $\mathbf{p} = \mathbf{p}_a$ (such that $m_p^2 > 0$) by this medium. In the general case, the quantity ΔU in Eq. (2) may determine, for example, a possible relative change in the energy of interaction between the electron and the medium upon emission of the Cherenkov photon. However, in this study, the term ΔU will be omitted.

Considering the phenomenon of Cherenkov radiation, "we are dealing essentially with radiation emitted by the medium under the action of the field of a moving particle" [12] (see also the statement from [6] cited above). Therefore, the corresponding Cherenkov photon possessing a finite mass m_p can accept part of the internal energy from the medium, as reflected in Eqs. (2) and (3) by the term ΔM which may sometimes (see Section 1) exactly coincide with m_p:

$$\Delta M = m_p = \frac{\varepsilon_p}{c^2}\sqrt{1 - \frac{v_p^2}{c^2}} = \begin{cases} \dfrac{\hbar\omega\sqrt{n^2 - 1}}{c^2 n}, & n > 1, \\ \dfrac{\hbar\omega\sqrt{1 - n^2}}{c^2}, & n < 1. \end{cases} \tag{4}$$

This relation, where m_p is a fixed characteristic of the medium only in the given case of coincidence with ΔM, determines the law of dispersion of the medium in the form of a dependence of n on ω for $n > 1$ and $n < 1$.

In the case of $m_p \ne \Delta M$ (for both $\Delta U = 0$ and $\Delta U \ne 0$ in Eq. (2)), the existence of a finite real m_p corresponding to the energy ε_p and the momentum $\mathbf{p}_a$ in a medium with $n \ne 1$ is insufficient for establishing a definite law of dispersion $n(\omega)$. Indeed, for this definition of m_p, the refractive index n can formally be even constant and independent of ω, although all real media encountered in the nature are, of course, dispersive, so that $n = n(\omega)$.

If, on the contrary, a dispersion law $n(\omega)$ is given, the m_p value in the general case (for $m_p \neq \Delta M$ in Eq. (4)) can also be a function of the frequency. In this case, m_p determines certain dispersion characteristics of the refractive medium. This quantity (as well as the refractive index $n(\omega)$) is not a relativistic invariant because it depends on the noninvariant frequency ω. For example, the observed optical frequencies of Cherenkov radiation at $n > 1$ correspond to m_p such that $m_p/m_e \sim 10^{-5}$. Note that, for a medium with $n \neq 1$, the frame related to the resting medium is always special [23]. Therefore, it is senseless to require the principle of equality of the inertial systems and the relativistic invariance to be observed in this system.

On the other hand, for $m_p = \Delta M$ (e.g., at $n < 1$), when $m_p = \Delta M = \hbar\omega_0/c^2$ (see Section 1), relation (4) may exactly correspond to the law of dispersion $n(\omega)$ for the transverse high-frequency electromagnetic waves in the isotropic collisionless plasma. In this case, the necessary change in the state of the medium upon emission of the Cherenkov photon corresponds, according to relation (4), to subtraction of a value equal to the plasmon energy from the total internal energy. Thus, emission of a "massive" Cherenkov photon by the medium at $n < 1$ and the threshold conditions (determined below) must be accompanied by the absorption of a condensate plasmon with arbitrarily small momentum $\mathbf{p}_0$, $\mathbf{p}_0$ corresponding, according to Eqs. (2) and (3) with $\Delta U = 0$, to the regime of $\mathbf{v}_2 \longrightarrow 0$, $M\mathbf{v}_2 \longrightarrow 0$ for $\Delta M \ll M$. In addition, relation (4) shows (exactly for $m_p = \Delta M$) that $\Delta M \longrightarrow 0$ for $\varepsilon_p \longrightarrow 0$ in Eq. (2) (with $\Delta U = 0$ and $\Gamma_0 \longrightarrow \Gamma_1$), which implies that the absence of changes in the energy state of the medium is unambiguously related to the absence of Cherenkov radiation. It is obvious that, for $\Delta M = m_p$, the value of ΔMc^2 determines according to relation (4) the lower boundary of the Cherenkov radiation frequency. Thus, ΔMc^2 can be considered as a part of the internal energy (in the form of "standing" elementary perturbations) that is potentially accessible for conversion into the energy of a running wave corresponding to Cherenkov radiation under over-threshold conditions determined below (see relation (5)) based on Eqs. (2)–(4).

Ryazanov [25] considered (from the context of Cherenkov radiation) an analogous possibility of emission of a high-frequency transverse quantum for the inverse Compton effect, that is, for the scattering of a high-energy electron on elementary Bose excitations in a medium (including a longitudinal plasmon) whereby this elementary excitation is absorbed by the electron [25], in the case of $n \longrightarrow 1$ and $|\mathbf{p}| \longrightarrow \varepsilon_p/c$ (like a photon in vacuum). The conservation equations (2) and (3) correspond to the case considered in [25], provided that we formally use $n = 1$, $\Delta U = 0$, and $M = 0$ and replace the terms $\Delta Mc^2\Gamma_2$ and $\Delta M\mathbf{v}_2\Gamma_2$ (transferred to the left-hand side of equations (2) and (3), respectively) by the energy $\varepsilon_p = \hbar\tilde{\omega}_0$ and the momentum $\mathbf{p}_0 = \hbar\mathbf{k}$ of the elementary excitation (in particular, longitudinal plasmon at $\tilde{\omega}_0 = \omega_0$) absorbed during the emission of Cherenkov radiation. In the real case of $M \neq 0$, but with $\mathbf{v}_2 \longrightarrow 0$, we obtain the same equations assuming that $M\mathbf{v}_2 \longrightarrow 0$ for all M. This limit corresponds (by virtue of $\Delta M \ll M$) to the long-wavelength limit $\mathbf{p}_0 \longrightarrow 0$ describing the condensate state of elementary Bose excitations of the medium.

On the other hand, it should be noted that system of equations (2), (3) differs from equations of the Ginzburg theory of Cherenkov radiation [10, 11] even by using the finite value of $\Delta M \neq 0$. Indeed, the former system coincides with the latter if we set $\Delta M = 0$ (for $\Delta U = 0$), reject all terms on the order of $O(v_2^2/c^2)$ in the limit of $\mathbf{v}_2 \longrightarrow 0$ in Eq. (2) (by putting $\Gamma_2 \longrightarrow 1$), and use in Eq. (3) the momentum $\mathbf{p}_c$ (the Abragam force impulse) instead of $\tilde{\mathbf{p}}_c = M\mathbf{v}_2\Gamma_2 \approx M\mathbf{v}_2$ when the aforementioned (see Section 1) relation between $\mathbf{p}_m$, $\mathbf{p}_c$, and $\mathbf{p}_a$ takes place:

$$\mathbf{p}_m = \mathbf{p}_c + \frac{\varepsilon_p \mathbf{v}_p}{c^2}.$$

However, it is obvious that the condition $\tilde{\mathbf{p}}_c = \mathbf{p}_c$ (implicitly used in the theory [10, 11]) is by no means valid in the general case, since even directions of the vectors $\mathbf{v}_2$ and $\mathbf{v}_p$ (or $\mathbf{k}$) for the Cherenkov photon emission do not always coincide. For example, in the case when the Cherenkov photon is emitted strictly in the direction of electron motion ($\cos\theta = 1$), the direction of recoil momentum $\tilde{\mathbf{p}}_c$ is opposite to the directions of vectors $\mathbf{k}$ and $\mathbf{p}_c$ (so that $\tilde{\mathbf{p}}_c \neq \mathbf{p}_c$).

In order to determine threshold conditions for the emission of Cherenkov radiation, let us analyze the relations (2)–(4), in particular, for $\mathbf{v}_2 \longrightarrow 0$ and $\tilde{\mathbf{p}}_c \longrightarrow 0$ (i.e., for the long-wavelength limit $\mathbf{p}_0 \longrightarrow 0$ of the elementary Bose perturbation absorbed during Cherenkov radiation), when $\Delta M = m_p$ in Eq. (4). If we divide both sides of Eq. 2 (with $\Delta U = 0$) by c, transferring all terms not related to the electron to the left sides of Eqs. (2) and (3), take the square of both sides of Eqs. (2) and (3), and subtract the result for Eq. (3) from that for Eq. (2) (this procedure is especially illustrative if performed using the 4-vector algebra), we arrive at the following condition for the emission of Cherenkov radiation for $n > 1$ (in the case of $n < 1$, n has to be everywhere replaced by $1/n$):

$$1 \geq \cos\theta = \frac{c}{v_0 n_*}\left(1 + \frac{\varepsilon\sqrt{n^2-1}}{\Gamma_0 n}\right), \qquad (5)$$

where $n_* = n + \sqrt{n^2-1}$ and $\varepsilon = \varepsilon_p/m_e c^2$. In particular, for $\varepsilon \ll 1$, Cherenkov radiation is possible for $v_0 > c/n_*$.

The structure of condition (5) is close to that in [11, Eq. (11)] (see also formula (1) in Section 1), but the two expressions only asymptotically coincide in the limit as $n \longrightarrow 1$. On the other hand, relations (1) and (5) qualitatively differ in that, first, formula (1) contains the negative quantity $\tilde{m}_p^2 < 0$, while condition (5) corresponds to the emission of a Cherenkov photon with nonzero and real effective mass ($m_p^2 > 0$) and, second (but even more significant), condition (5) contains information about the condensate ($\mathbf{p}_0 \longrightarrow 0$) elementary Bose excitation with the energy $\varepsilon_0 = \Delta M c^2 \neq 0$ (coinciding with $m_p c^2$ for (4) and (5)) absorbed in the medium during the emission of Cherenkov radiation. As will be shown in the next section based on the relativistic generalization of the Landau criterion, it is the latter circumstance that makes the emission of Cherenkov photons energetically favorable under condition (5).

The condition for the Cherenkov radiation emission at $n < 1$ (e.g., in the case when isotropic plasma is the refractive medium) can be obtained from Eq. (5) by substituting $1/n$ for n (in this case, $n_* > 1$). Indeed, for $n < 1$

$$n_* = \frac{n}{1 - \sqrt{1 - n^2}} > 1.$$

This is the basic difference between condition (5) and conclusions of the classical theory [5–11] for the emission of Cherenkov photons, since the former condition admits direct and effective nonbremsstrahlung Cherenkov radiation emission from collisionless isotropic plasma in the form of transverse high-frequency electromagnetic waves. Tamm [8] believed that only longitudinal plasma waves could be emitted for $n < 1$ in the case when the electron velocity is greater then the phase velocity of such waves.

2.2. Relation (5) and its analog for $n < 1$ determine the possibility of Cherenkov photon emission only provided that $m_p = \Delta M$ in formula (4). In the case of $m_p \neq \Delta M$ and a medium with arbitrary dispersion (not determined from (4) at a fixed m_p value), it is possible to obtain from Eqs. (2) and (3) a more general condition instead of (5) for the Cherenkov radiation emitted upon absorption of an ultimately long-wavelength ($\mathbf{p}_0 \longrightarrow 0$) elementary Bose excitation of a medium at an energy of $\varepsilon_0 = \hbar\tilde{\omega}_0$. For example, at $n > 1$, condition (5) is replaced by

$$1 \geq \cos\theta = \frac{cn}{v_0}\left[y - \frac{\varepsilon}{2\Gamma_0}\left(y^2 - \frac{1}{n^2}\right)\right] \geq -1, \qquad (6)$$

where $y = 1 - \varepsilon_0/\varepsilon_p$. For $\varepsilon_0 = m_p c^2$ (i.e., for $m_p = \Delta M$ in formula (4)), condition (6) completely coincides with (5). In the case of $n < 1$, the desired condition is obtained by substituting $1/n$ for n in relation (6).

In contrast to threshold condition (5), whose validity was related to the finiteness of m_p (and ΔM, since this relation was obtained for $\Delta M = m_p$), relation (6) at $n_n \neq 1$ admits the emission of Cherenkov photons even in the limit of $n \longrightarrow 1$ (when $m_p \longrightarrow 0$, but $\Delta M = \varepsilon_0/c^2 \neq 0$). Note that this case also includes the conditions considered in [25] (although no threshold relations of the type of (6) were reported in that paper). The emission in [25] was studied, in fact, for $n = 1$ (in the description of the photon momentum), which is inconsistent with the very presence of the medium (nevertheless implied by assuming the existence of elementary excitations). Therefore, it would be more correct to consider the limit as $n \longrightarrow 1$, rather than the case when $n = 1$. Cherenkov radiation was not even mentioned in [25], where neither the possibility of anisotropic emission of a transverse photon upon absorption of the elementary excitation by the most energetic particle was considered, nor any thresholds of the particle velocity were established in contrast to what is characteristic of nonbremsstrahlung Cherenkov photon emission under condition (6).

For example, in the case of $\varepsilon \ll 1$ and $y \approx O(1)$, inequalities (6) describing conditions for the possible existence of Cherenkov radiation can take place only for the following limitations imposed on the electron velocity and the frequency of emitted radiation quanta:

$$\begin{aligned} c > v_0 > cn\left|1 - \frac{\tilde{\omega}_0}{\omega}\right|, \\ \frac{n\tilde{\omega}_0}{n+1} < \omega < \frac{\tilde{\omega}_0 n}{n-1}, \quad n > 1 \end{aligned} \qquad (7)$$

(for $n < 1$, n has to be replaced by $1/n$). In particular, $\tilde{\omega}_0 = \omega_0$ corresponds to the case when the emission of Cherenkov radiation from an isotropic plasma for $n < 1$ requires absorption of the condensate ($\mathbf{p}_0 \longrightarrow 0$) plasmon. For $\hbar\omega_0 \neq m_p c^2$, relations (7) in contrast to condition (5), not only lead to different forms of the limitations with respect to velocity v_0, but imply additional limitations with respect to the radiation frequency ω (which follows from the condition of consistency of inequalities (7) for v_0). In the limit of $n \longrightarrow 1$, this additional condition takes the form of $\omega_0/2 < \omega$. This limitation with respect to ω following from inequalities (7) will be used below for comparison with the results of solar radio burst observations.

Conditions (7) imply that Cherenkov radiation can be emitted not only for $\omega > \tilde{\omega}_0$, when the electron slows down as a result of the Cherenkov photon emission. Emission is also possible for $\omega < \tilde{\omega}_0$, when the energy of absorbed elementary excitation is partly spent for acceleration of the electron. This mechanism of the charged particle acceleration at the expense of Cherenkov radiation can be realized in a turbulent plasma,

were the values of $\tilde{\omega}_0$ can be sufficiently large. In contrast to the well-known mechanisms [29–31] of particle acceleration in cosmic rays, this self-acceleration of a charged particle is due to the very possibility of Cherenkov radiation emission under the condition that the initial velocity of the particle exceeds the threshold determined by relations (7). This threshold velocity can be rather small, provided that $\omega \longrightarrow \tilde{\omega}_0$ (i.e., that the Cherenkov radiation frequency is close to the intrinsic frequency $\tilde{\omega}_0$ of absorbed elementary excitation).

It is evident that condition (6) for the emission of Cherenkov radiation is satisfied not only in the case when limitations (7) are obeyed (which correspond to $\varepsilon \ll 1$, that is, to not very high frequencies of Cherenkov radiation quanta). This condition can be also valid for a hard Cherenkov radiation in the γ and X-ray ranges, where $\varepsilon \approx O(1)$ or even $\varepsilon \gg 1$. This case obviously corresponds to $n < 1$ and, hence, condition (6) has to be modified by substituting $1/n$ for n.

In the case of $y < 0$ (which corresponds to the acceleration of electron upon Cherenkov photon emission), general conditions (6) are satisfied for $|y| < n < 1$ (that is, for $\varepsilon_p < \varepsilon_0 < \varepsilon_p(n + 1)$), while for $n > 1$ these conditions are valid only when $|y| < 1/n$ (or $\varepsilon_p < \varepsilon_0 < \varepsilon_p(n + 1)/n$). Therefore, the acceleration is negligibly small for $n \gg 1$ and $n \ll 1$, but for $n \longrightarrow 1$ it can be finite and increases with ε_p (since $\varepsilon_0 - \varepsilon_p \leq \varepsilon_p n$ for $n \longrightarrow 1$). The corresponding limitation imposed on v_0 (following from condition (6) for $y < 0$ and $n < 1$) is presented in the Appendix (see Eq. (A.1)).

On the other hand, for $y > 0$ we obtain a threshold condition (which also follows from relations (6)) with respect to v_0 for Cherenkov photon emission. For example, in the case of $n < 1$ and $y < n$, this condition appears as

$$v_0 > c\beta_1,$$

$$\beta_1 = \frac{ny + \frac{\varepsilon}{2}(n^2 - y^2)^{3/2}\left[1 + \frac{\varepsilon^2}{4}(n^2 - y^2)\right]^{1/2}}{n^2 + \frac{\varepsilon^2}{4}(n^2 - y^2)^2}, \quad (8)$$

where $\beta_1 < 1$ for any ε.

It should be noted that, in the limit of $\varepsilon_0 \longrightarrow 0$ (i.e., for $y \longrightarrow 1$), condition (6) exactly coincides with relation (1) in which the negative quantity $\tilde{m}_p^2$ is formally replaced by $m_p^2 > 0$ expressed via $\mathbf{p}_a$. In this case, for m_p not necessarily coinciding with $\Delta M = \varepsilon_0/c^2$, we obtain a condition under which the electron directly (without absorption of an elementary excitation in the medium) emits a photon with a nonzero real mass m_p.[1]

[1] If $\tilde{m}_p^2$ in relation (1) is replaced by m_p^2, this condition can be satisfied for $n > 1$ only at $\varepsilon > 2\Gamma_0 n/(n + 1)$).

As will be shown based on the relativistic generalization of the Landau criterion obtained below, this process of emission directly from the electron is not energetically favorable, although conditions (6) are formally satisfied for $\varepsilon_0 = 0$. Moreover, the case of $y > n$ at $n < 1$ also does not correspond to energetically favorable emission despite the fact that conditions (6) are met in the form of limitations (not much more complicated than relation (8)) with respect to v_0.

2.3. Now let us consider in more detail the threshold conditions for Cherenkov photon emission by comparing relation (6) for $n \longrightarrow 1$ to the conditions established in [25] for the emission of massless hard transverse quanta by a high-energy particle upon absorption of an ultimately long-wavelength elementary excitation of a medium. With a view to the particular case (also studied in [25]) when this excitation has the form of a longitudinal plasmon, we can represent for $n < 1$ the following limitation with respect to the frequency of emitted quanta ensuring the validity of conditions (6) for the emission of Cherenkov photons in the limit of $n \longrightarrow 1$, $\varepsilon(1 - n^2) \approx o(1)$, and $\varepsilon \approx O(1)$:

$$\varepsilon_2 < \frac{\omega}{\omega_0} < \varepsilon_1, \quad (9)$$

where

$$\varepsilon_{1,2} = \frac{1 + \tilde{\varepsilon}\gamma_0/2}{1 + \tilde{\varepsilon}\gamma_0 \mp \beta n}, \quad \tilde{\varepsilon} = \frac{\hbar\omega_0}{m_e c^2},$$

$$\beta = \frac{v_0}{c}, \quad \gamma_0 = \sqrt{1 - \beta^2} = \frac{1}{\Gamma_0}.$$

A condition for photon emission that was implicitly used (although not explicitly formulated) in [25] for $n = 1$, corresponds to the following inequality, which can be derived from Eqs. (2) and (3):

$$|\Delta\mathbf{p}_0|_{\min} = -\frac{|\Delta\mathbf{p}_0|_{\min}}{m_e c}\gamma_0 = \beta - \varepsilon n\gamma_0 + -\sqrt{(1 - \varepsilon y\gamma_0)^2 - \gamma_0^2} < 0, \quad (10)$$

where $\varepsilon y \equiv \varepsilon - \tilde{\varepsilon}$ and $n < 1$ (for $n > 1$, n has to be replaced by $1/n$).

Inequality (10) determines the condition for absorption of an elementary excitation of a medium with an arbitrarily small momentum $\mathbf{p}_0$, taking into account that a minimum momentum lost by a quasi-particle upon absorption corresponds to the case when a photon is emitted in the direction of motion of a high-energy particle [25], that is, to $\theta = 0$. For $y > 0$, inequality (10) should be supplemented by the limitation $\gamma_0 y < 1$, following from the energy conservation law, and by the condition $\varepsilon\gamma_0 y < 1 - \gamma_0$ expressing positive definiteness

of the radicand in (10). Relation (10) has physical sense only provided that all these conditions are satisfied. However, these conditions were not considered at all in [25]. Note that, for example, in the case of $\varepsilon \geq \beta/n\gamma_0$, when inequality (10) is automatically satisfied, it is especially important to observe the aforementioned supplementary conditions. On the contrary, it is only for $\varepsilon < \beta/n\gamma_0$ that relation (10) at a fixed γ_0 leads to a stronger limitation (more restrictive than the two additional conditions),

$$\varepsilon\gamma_0 y < 1 - \sqrt{\gamma_0^2 + (\beta - n\varepsilon\gamma_0)^2},$$

from which we obtain the exact threshold (8) with respect to v_0 (e.g., for $0 \leq y < n \leq 1$).

A condition for the onset of emission presented in [25] (see [25, Eq. (1.3)]) can be also obtained from inequality (10) for $n = 1$ and in the limit of $\gamma_0 \ll 1$ under the additional conditions that $\gamma_0 y\varepsilon \approx O(1)$ and $\tilde{\varepsilon} \leq O(\gamma_0)$ (not explicitly presented in [25]). The condition [25, Eq. (1.3)] can be written as

$$\frac{\omega}{\omega_0} < \frac{2}{\gamma_0^2(1 + 2\tilde{\varepsilon}/\gamma_0)}, \qquad (11)$$

which corresponds to the emission of radiation with a frequency $\omega \gg \omega_0$ and agrees with the necessary restrictions for ε and $\tilde{\varepsilon}$ indicated above, for which relation (11) is valid (i.e., $\varepsilon \approx O(1/\gamma_0)$ and $\tilde{\varepsilon} \leq O(\gamma_0)$ for $\gamma_0 \ll 1$. For these very limitations with respect to ε and $\tilde{\varepsilon}$ (not indicated in [25]), inequality (11) asymptotically coincides in the limit of $\gamma_0 \ll 1$ with the right-hand side of inequality (9) determining the restrictions necessary for the validity of threshold condition (6) for the emission of Cherenkov photons. For $\gamma_0 \ll 1$ and $n = 1$, the left-hand side of inequality (9) corresponds to $\omega > \omega_0/2$ and, hence, is automatically satisfied under condition (11) for which it is required that $\omega \gg \omega_0$.

Thus, the condition of hard photon emission formulated in [25] asymptotically coincides (in the limit considered in that study) only for $n = 1$ with the conditions (9) necessary for obeying the threshold condition (6) for the emission of asymmetric nonbremsstrahlung Cherenkov radiation. The asymmetric character of radiation was not mentioned in [25], since the scope of that study did not include the phenomenon of Cherenkov radiation because of the high frequency character of the transverse emission (traditionally excluded for $n < 1$ from the consideration within the framework of the classical theory of Cherenkov radiation [5–11]). For this reason, the problem of determining the threshold velocity of a charged particle, above which such nonbremsstrahlung emission of a hard transverse quantum is possible upon absorption of an elementary excitation with even a relatively small energy, was not considered in [25]. Using condition (11) with an additional assumption that $\tilde{\varepsilon}/\gamma_0$ is ultimately small ($\tilde{\varepsilon}/\gamma_0 \ll 1$), it is easy to obtain a threshold condition for the velocity v_0, which exactly coincides with inequalities (7) for $n \longrightarrow 1$ and $\tilde{\omega}_0 = \omega_0$. The corresponding condition for the ratio of frequencies ω and ω_0 in (7) is also observed, since condition (11) corresponds to the limit of $\omega \gg \omega_0$ in which the left-hand side of inequality (7) for the frequencies is satisfied, while the right-hand side is met exactly for $n \longrightarrow 1$, when $\omega_0/(1 - n) > \omega$ at $n < 1$.

It should be noted that Gailitis and Tsytovich [26] considered (in contrast to [25]) the case of relatively short-wavelength longitudinal plasmons transformed as a result of the inverse Compton effect on a relativistic electron into transverse electromagnetic waves. The energy and momentum equations used in [26] can be also obtained from Eqs. (2) and (3) by taking $\Delta U = 0$, $n = 1$ (i.e., $|\mathbf{p}| = \varepsilon_p/c$), and $M = 0$, but (in contrast to [25]) for finite velocities $\mathbf{v}_2$ corresponding to the transverse plasmon energy $\hbar\omega_0 = \Delta Mc^2\Gamma_2$ and a finite (rather large) momentum $\mathbf{p}_0 = \Delta M\mathbf{v}_2\Gamma_2 = \hbar\tilde{\mathbf{k}}$ (for which the corresponding phase velocity is small compared to c in accordance with the limit of $\omega_0/\tilde{k} \ll c$ considered in [26]). The conclusions obtained in [26], including those concerning the mechanism of emission, are related to the use of this short-wavelength limit with respect to $\tilde{k}$ (in contrast to [25] and this study).

3. GENERALIZATION OF THE LANDAU CRITERION

This section presents relativistic generalization of the Landau criterion [20] used to establish conditions under which the process of Cherenkov photon emission in a system comprising a medium and a sufficiently fast charged particle (electron) is an energetically favorable process analogous to the anomalous Doppler effect or dissipative instability [11, 22]. Similar to the preceding section, we will first consider the case when the energy $\varepsilon_0 = \Delta Mc^2$ of the elementary excitation absorbed during Cherenkov photon emission exactly coincides with m_pc^2 according to (4) (e.g., for $\varepsilon_0 = \hbar\omega_0$) and then proceed to the case of arbitrary (not necessarily related to m_p) values of ΔM and ε_0.

3.1. In order to obtain a relativistic generalization of the Landau criterion combining the phenomenon of Cherenkov radiation and anomalous Doppler effect, let us follow Landau [20] and consider a representation of the energy balance equation (2) (for $\Delta U = 0$, $v_2 \longrightarrow 0$, $\Gamma_2 \longrightarrow 1$, and with allowance for relation (4)) using the frame of reference moving at a constant velocity $\mathbf{v}_0$,

where the electron was at rest before emission of the Cherenkov photon. Using Eqs. (2) and (4), we obtain

$$m_e c^2\left[1-\Gamma_0\Gamma_1\left(1-\frac{\mathbf{v}_0\cdot\mathbf{v}_1}{c^2}\right)\right] = \Gamma_0 Q, \qquad (12)$$

where

$$Q = Q_1 = \varepsilon_p\left[1-\frac{\mathbf{v}_0\cdot\mathbf{v}_p}{c^2}-\sqrt{1-\frac{v_p^2}{c^2}}\right].$$

Here, we use the formula $\Delta M c^2 = \varepsilon_p\sqrt{1-v_p^2/c^2}$ obtained from relation (4) and corresponding to the value m_p fixed as $m_p = \Delta M$.

Since the left-hand side of Eq. (12) is negative for any $\mathbf{v}_1 \neq \mathbf{v}_0$ and is zero for $\mathbf{v}_1 = \mathbf{v}_0$, a condition of the Cherenkov photon emission (with $\varepsilon_p > 0$) generalizing the Landau criterion is $Q < 0$ in Eq. (12). The inequality $Q < 0$ yields a condition for the energetically favorable emission of Cherenkov photons (accompanied by a negative change in the energy of the medium, see [20, 32]):

$$\cos\theta > c/v_0 n_*, \qquad (13)$$

where n_* was determined above in relation to (5) for both $n > 1$ and $n < 1$. Using condition (13) and taking into account that $|\cos\theta| \leq 1$, we obtain the inequality determining a threshold velocity for the Cherenkov photon emission:

$$v_0 > v_{\text{th}} = c/n_*. \qquad (14)$$

Note that, in the nonrelativistic case ($v_0 \ll c$, $v_p \ll c$), the condition $Q < 0$ in Eq. (12) for $\varepsilon_p > 0$ exactly coincides with the well-known Landay criterion [20, 32],

$$E - \mathbf{p}\cdot\mathbf{v}_0 < 0, \qquad (15)$$

determining the possibility of energetically favorable creation of excitations with the energy E and a momentum $\mathbf{p}$, in a medium moving at a constant velocity of $-\mathbf{v}_0$ in the indicated system of frame. In relation (15), the energy E and the momentum $\mathbf{p}$ are not mutually related like ε_p and $\mathbf{p}$, because E (in contrast to ε_p) is not equal to the total excitation energy (that includes the internal energy determined by the rest mass of a particle of the medium carrying the excitation). In the limit of $v_p \ll c$ (where v_p is the velocity of propagation of the excitation energy flux), the energies E and ε_p are related as

$$E = \varepsilon_p\left(1-\sqrt{1-\frac{v_p^2}{c^2}}\right) \approx \frac{\varepsilon_p v_p^2}{2c^2}.$$

Using criterion (15) and taking into account that $|\mathbf{p}| = \varepsilon_p v_p/c^2$ and $E = v_p p/2$, we obtain a nonrelativistic analog of the threshold condition (14):

$$v_0 > \frac{v_p}{2} = \frac{E(p)}{p},$$

where $v_p/2 \equiv \partial E/\partial p = v_g$. For this definition of the threshold velocity, we observe the equality (pointed out in [20] of the group and phase velocities of elementary excitations corresponding to their creation [20, 32]. According to the classical theory of Cherenkov radiation [9], the phase and group velocities of light in a transparent medium also coincide for the threshold conditions of Cherenkov photon emission.

3.2. Now let us proceed to the relativistic generalization of the Landau criterion in the general case, when both $\Delta M \neq m_p$ and $\Delta M = m_p$ is possible in relation (4). In this case, the energy $\varepsilon_p = \hbar\omega$ of the emitted Cherenkov photon is not always (only for $\Delta M = m_p$) uniquely related via formula (4) to the energy $\varepsilon_0 = \hbar\tilde{\omega}_0$ of the Bose excitation absorbed during Cherenkov photon emission. No such relation takes place even for $\tilde{\omega}_0 = \omega_0$, although (as was noted above) this case at $n < 1$ is most adequate to the relation (4) with $\Delta M = m_p$. In the right-hand side of Eq. (12), we obtain a different representation for Q:

$$Q = Q_2 = \varepsilon_p\left(1-\frac{\mathbf{v}_0\cdot\mathbf{v}_p}{c^2}\right)-\varepsilon_0.$$

In the case of emission with $\varepsilon_p > 0$, we obtain the following condition necessary for the quantity Q to be negative:

$$\cos\theta > \frac{yn}{\beta}, \qquad (16)$$

where $n > 1$ (for $n < 1$, n has to be replaced by $1/n$). Since $|\cos\theta| \leq 1$ in relation (16), the case of $y > 0$ implies a threshold with respect to the velocity v_0. This limitation, following from the inequality $yn/\beta < 1$, coincides with condition (7) not only for $\varepsilon \ll 1$ and $y = O(1)$ (as in the case when condition (7) was derived from relation (6)). For $y < 0$, we obtain either limitation for v_0 (this case, exactly corresponding to inequalities (7), reveals the anisotropic cone of Cherenkov radiation according to condition (16)) or the condition $|y|n/\beta > 1$. In the latter case, condition (16) reduces to inequality $\cos\theta \geq -1$, not necessarily determining any anisotropy in the emission. However, the anisotropy of Cherenkov radiation can still be determined by condition (6) (see Appendix, Eq. (A.1)). Thus, it is obvious that the theory of Cherenkov radiation must take into account the finite energy ε_0 lost by the medium. This is necessary to provide for the energetically favorable emission of Cheren-

Table

Medium	n	n_*	$\cos\theta_m^A$	$\cos\theta_m^B$	β_*^A	β_*^B	β^A	β^B
H_2O	1.3371	2.2247	0.6691	0.7431	0.6718	0.6049	1.1177	1.0064
C_6H_{12}	1.4367	2.4683	0.5	0.6428	0.8103	0.6303	1.392	1.083
C_6H_6	1.5133	2.6491	0.454	0.5736	0.8315	0.6581	1.4556	1.152
$C_{11}H_{12}O_2$	1.5804	2.8049	0.3584	0.5	0.9519	0.6217	1.689	1.103

kov photons, since the right-hand part of inequality (16) can be smaller than unity only provided that $\varepsilon_0 \neq 0$ (for both $n > 1$ and $n < 1$). In this respect, a useful example is offered by the case of $\varepsilon_0 = 0$ considered above (with the substitution $\tilde{m}_p \longrightarrow m_p$ in Eq. (1)), in which the necessary conditions of Cherenkov photon emission are satisfied, but this process is not energetically favorable and, hence, not necessarily realized.

With respect to the conditions of emission following from inequality (10), the obtained relativistic generalization (16) of the Landau criterion too only refines limitations necessary for the energetically favorable realization of Cherenkov radiation (see Appendix). The necessary conditions (10) actually coincide with the limitation with respect to v_0 following from (16) only in the limit of $z \ll 1$ ($z = \varepsilon/n$ for $n > 1$; $z = \varepsilon n$ for $n < 1$) and $y \approx O(1)$. In other cases, relation (10) leads to limitations of the type of (8), which are stronger than condition (16). However, only condition (16) separates an important subclass of limitations determined by inequality (10), which are necessary to ensure energetically favorable realizations of Cherenkov radiation.

Thus, the threshold conditions for v_0 following from the obtained relativistic generalization of the Landau criterion determine the necessary (although not always sufficient) conditions for energetically favorable realizations of Cherenkov radiation. However, based on an analysis of the necessary threshold conditions established using this approach, it is already possible to understand the basic physical mechanism of nonbremsstrahlung emission of Cherenkov radiation by the medium, as energetically favorable production of a Cherenkov photon with simultaneous absorption of a condensate Bose excitation in the medium, caused by its interaction with a sufficiently fast charged particle (electron). In the frame of reference corresponding to the energy equation representation in the form of (12) (for $Q = Q_1$ or $Q = Q_2$), the emission of a Cherenkov photon with energy $\varepsilon_p > 0$ under condition (13) or even (16) for $\varepsilon_0 < \varepsilon_p$ at $y > 0$ leads to a simultaneous increase in the kinetic energy of the charged particle (provided that $\mathbf{v}_0 \neq \mathbf{v}$), which is analogous to the anomalous Doppler effect or the realization of "negative energy" in various physical systems [11, 22, 24, 33]. This mechanism of Cherenkov radiation emission in fact offers an example of dissipative instability [21–24], whereby it is energetically favorable for a system in the over-threshold regime to create perturbations realized, in this case, in the form of Cherenkov photons.

4. CHERENKOV RADIATION THEORY AND EXPERIMENTAL DATA

In this section, we will compare (Section 4.1) conclusions of the proposed theory of to some experimental data (see [3, Ref. 2]) and the classical Tamm–Frank–Ginzburg theory [5–11]. Then, we will briefly consider (Section 4.4) the possibility of applying the results to interpretation of some radio astronomy data.

4.1. Experimental data on the angular distribution of Cherenkov radiation intensity [3, Fig. 4] observed in four liquid transparent media, treated in terms of condition (13) and the classical theory [5–11], yield the results summarized in the table, where superscript A refers to the data obtained using a ThC" γ-radiation source and superscript B, to the results for Cherenkov radiation excited by Compton electrons generated by γ radiation from a Ra source. According to [3], the emission was observed in the entire range of angles $0 \leq \theta \leq \theta_m^{A,B}$, with a maximum intensity at $\theta = \theta_0^{A,B} < \theta_m^{A,B}$, a sharp threshold at $\theta = \theta_m^{A,B}$, and the absence of emission at $\theta > \theta_m^{A,B}$. The emission intensity at $\theta = 0$ was lower than that at the maximum ($\theta = \theta_0^{A,B}$) only by a factor of 1,5–3. The values of $\beta_*^{A,B}$ were calculated using relation (13) determining a necessary threshold for the energetically favorable realization of Cherenkov radiation,

$$\beta_*^{A,B} = \frac{1}{n_* \cos\theta_m^{A,B}}.$$

The quantities $\beta^{A,B}$ were determined according to the classical theory [5–11], where the threshold velocity v_0 corresponds to

$$\beta^{A,B} = \frac{1}{n \cos\theta_m^{A,B}}.$$

The calculations were performed using experimental data on the threshold angles $\theta_m^{A,B}$ observed for Cherenkov radiation generated in various media (see table) using the aforementioned excitation sources.

Note that the values of $\beta_*^{A,B}$ and $\beta^{A,B}$ correspond to the threshold values of v_0/c, which must be always smaller than unity. Therefore, the values of $\beta^{A,B}$ determined using the classical theory [5–11] obviously fail to meet this requirement of the relativistic theory, while the values of $\beta_*^{A,B}$ calculated according to the proposed theory obey the condition. However, replacing the limiting angles $\theta_m^{A,B}$ by the values ($\theta_0^{A,B}$) corresponding to the observed interference maximum in the Cherenkov radiation intensity distribution, we obtain $\beta^{A,B}$ values satisfying the requirement of $\beta^{A,B}(\theta_0^{A,B}) < 1$ for the data from [3].

Thus, a comparative analysis of theoretical conclusions and the experimental data for Cherenkov radiation suggests that the necessary condition (13) (or (16)) allows us to determine whether an energetically favorable realization of Cherenkov radiation is possible in the entire range of θ angles corresponding to this threshold. At the same time, criteria based on the stronger inequalities (5), (6), and (10) may also determine the corresponding sufficient conditions for the emission of Cherenkov photons, provided there is consistency with conditions (13) or (16).

The classical theory of Cherenkov radiation [5–11] also poses more rigid restrictions. A comparison to experimental data shows that this theory does not elucidate the energy mechanism of the threshold emission of nonbremsstrahlung Cherenkov radiation and corresponds to description of a coherent radiation existing in the medium, in good agreement with the observed interference maxima of the Cherenkov radiation intensity.

Of course, the above comparison to experimental data [3] is by no means exhaustive. In particular, we did not take into consideration scatter (always existing in a real experiment) in the directions of motion of electrons initiating the emission of Cherenkov photons by the medium, which may result in broadening of the observed intensity distributions $I(\theta)$, in particular, relative to the $\theta = \theta_0$ direction. It should be noted that, using a modified classical theory of Cherenkov radiation [6], which takes into account a finite time τ during which an electron radiates while moving at a sufficiently high constant velocity, it is also possible to describe a finite broadening of the intensity distributions $I(\theta)$ to within $\Delta\theta$ near $\theta = \theta_0$ (for $\cos\theta_0 = 1/\beta n$, $n > 1$), where

$$\Delta\theta \approx \frac{1}{\omega\tau(\beta^2 n^2 - 1)^{1/2}}.$$

However, this $\Delta\theta$ value decreases with increasing n, which contradicts the observed [3] growth of $\Delta\theta$ with n. Moreover, an increase in $\Delta\theta$ according to the modified theory is symmetric relative to $\theta = \theta_0$ and $\theta = -\theta_0$, which is also at variance with the results of observations [3]: the broadening of $I(\theta)$ in this angular region exhibits a significant asymmetry, being much more pronounced for $\theta < |\theta_0|$ than for $\theta > |\theta_0|$ ($I(\theta) = 0$ for $\theta \geq \theta_m$). Indeed, symmetric (with respect to θ_0) extrapolation to $\Delta\theta = \theta_m - \theta_0$ must (according to the data for $\Delta\theta$ [3]) give $I(\theta) = 0$ for $\theta = 0$. However, this is not observed in [3], where $I(0) \geq I(\theta_0)/3$ and $I(\theta_0) = \max I(\theta)$ (the intensity of Cherenkov radiation corresponding to the interference maximum).

Thus, using conclusions based on the relativistic generalization of the Landau criterion, it is possible to determine the possibility of energetically favorable realization of Cherenkov radiation in the entire range of θ angles (determined from relation (13) for $m_p = \Delta M$ or from condition (16) for an arbitrary relation between m_p and ΔM), rather than in a direction fixed by conditions (6) or (5), in which the emission is realized only for $\mathbf{v}_0 = \mathbf{v}_1$ (when the left-hand side of Eq. (12) is zero). The above interpretation of the mechanism of nonbremsstrahlung Cherenkov photon emission is valid (in the frame of reference for which the energy equation (12) was written) even without taking into account the internal degrees of freedom of the charged particle. The presence of such degrees of freedom, or explicit allowance in relations (2) and (12) for the heat evolution or some other manifestations of the dissipative interaction between the particle and medium during Cherenkov photon emission implies the need for introducing the corresponding additional terms of the ΔU type. This does not change the obtained necessary conditions (13) and (16) for realization of the energetically favorable Cherenkov radiation; on the other hand, this cannot be adequately reflected in the momentum balance equation. For this reason, unambiguous determination of the angle of Cherenkov radiation emission using conditions such as (5) or (6) becomes impossible. In deriving the nonrelativistic Landau criterion determining creation of the vortex ("roton") modes in superfluid helium, the analysis was also based on the energy balance equation (without recourse to the momentum balance equation) [20, 32].

4.2. It should be noted that the approach developed in this study, which takes into account a finite rest mass related to the created elementary excitation, may be useful for refinement of the nonrelativistic theory [20, 32] by including into consideration some factors determining the formation of this effective mass, such as the system geometry and boundary conditions. Indeed, experimentally observed values of the critical helium flow velocity [34] have proved to be significantly lower than theoretical predictions [20]. Note that the expression obtained above for the critical velocity in the nonrelativistic limit is half as small as the excitation prop-

agation velocity v_p determining the flux of the total energy ε_p.

4.3. Separate consideration is required for elucidation of the specific physical mechanisms involved in the formation of condensate elementary Bose excitations of a given medium, which can be converted into Cherenkov photons in cases of realization of the energetically favorable threshold emission of nonbremsstrahlung Cherenkov radiation by this medium. Here, it will be only briefly mentioned that it is possible to determine ΔM (or $\varepsilon_0 = \hbar\tilde{\omega}_0$) via the effective electromagnetic rest mass $m_{em} \approx e^2/n^2c^2R$ of an electron in a medium with $n \neq 1$, which accounts for the electromagnetic impulse of an electron moving at a constant velocity (in [35, Eq. (28.7)], m_{em} was introduced in the case of vacuum with $n = 1$). In order to estimate the effective polarization radius R of an electron, let us assume that $m_{em} \approx \Delta M$ and use the above estimate $\Delta M = m_p \sim 10^{-5}m_e$. As a result, we obtain

$$R \approx \frac{10^5 e^2}{n^2 m_e c^2},$$

which implies that the formation of an elementary Bose excitation absorbed in a medium during Cherenkov radiation emission may involve a volume of this medium with the size comparable (for $n \approx O(1)$) to atomic dimensions: $R \sim 10^{-8}$ cm. The finite electron velocity v_0 determines the corresponding electromagnetic impulse in the medium. However, for $v_0 < v_{th}$, the field cannot "break away" (in the form of a running wave) from the corresponding volume of the medium near which the electron moves at a given moment. For $v_0 > v_{th}$ (when, according to conditions (13) and (16), creation of a Cherenkov photon by the medium becomes energetically favorable), a standing wave (existing at $v_0 < v_{th}$ within certain effective resonator of size R) is capable of transforming into a running wave observed in the form of Cherenkov radiation. The minimum frequency of this running wave formed at $v_0 > v_{th}$ is determined by the intrinsic frequency $\tilde{\omega}_0$ (or by the value of $\Delta M = \hbar\tilde{\omega}_0/c^2$) of this standing (at $v_0 < v_{th}$) wave (see [35, p. 233]).

In particular, as was noted above, the elementary excitation absorbed during Cherenkov photon emission at $\tilde{\omega}_0 = \omega_0$ and $n < 1$ represents the well-known condensate longitudinal plasmon. In other cases, for example, in determining $\tilde{\omega}_0$ from the energy of thermal excitations of the medium (when $\tilde{\omega}_0 \approx kT/\hbar$), it is more difficult to define physical conditions for the formation of the corresponding quasi-particles of the medium capable of potentially transforming into Cherenkov photons. The above example of determining the effective polarization radius of electron may serve as a guide in approaching such problems, especially for media with $n > 1$.

4.4. The theory of nonbremsstrahlung Cherenkov radiation developed in this study refines the physical mechanism of this phenomenon and admits more adequate and wider application than the classical theory [5–11]. In particular, the proposed theory can be used for the interpretation of many astrophysical phenomena, only some of which are briefly mentioned below.

At present, the origin of solar radio bursts of type III is explained in [27] in terms of the classical theory of Cherenkov radiation [5–11], according to which a beam of sufficiently fast electrons can induce the emission of only longitudinal plasma waves, only a small part of which are capable of converting into transverse plasmons upon interaction with the beam. As was noted above, this conversion was discussed, for example, in [25, 26] where various (although formally similar) mechanisms of emission of transverse quanta were considered. It was pointed out [25, 26] that such conversion provides a more effective generation of the transverse electromagnetic waves than, for example, the magnetic bremsstrahlung mechanism of emission.

As was shown in this paper, the mechanism of emission considered in [25] is in fact a particular case of nonbremsstrahlung Cherenkov photon emission, whose energetically favorable realization is accompanied by the absorption of a condensate (possessing arbitrarily small momentum) longitudinal plasmon. Moreover, the theory of Cherenkov radiation developed in this study admits (in a medium with $n < 1$) direct (not involving preliminary generation of longitudinal plasmons by an electron beam) threshold (with respect to the electron velocity) generation of transverse electromagnetic waves by the medium, which can also provide a more effective generation, for example, of type III solar radio bursts.

There are also necessary (but not always sufficient) grounds for a natural explanation of the observed excess frequency of type III solar radio bursts as compared to the plasma frequency in the region from which radiation is outgoing. Various hypotheses have been suggested for explaining this phenomenon, from unconvincing assumptions of doubled plasma density (as compared to that at the corresponding altitude in the real corona) to more reasonable considerations involving refraction and scattering effects [27]. At the same time, limiting condition (7) for the Cherenkov radiation frequency ω (for $n < 1$ and $\tilde{\omega}_0$ coinciding with the plasma frequency ω_0, that is, in the case of a condensate longitudinal plasmon absorbed during Cherenkov photon emission) implies the possibility of radiating not only at $\omega = \omega_0$, but at $\omega \approx \sqrt{2}\,\omega_0$ as well. For example (7), in the case of $n \approx 0.5$ this limiting condition indicates that $2\omega_0/3 < \omega < 2\omega_0$. According to relation (7), at $n < 1$, the

relative width of the Cherenkov radiation frequency band

$$\delta = \frac{\Delta\omega}{\omega_0} = \frac{2n}{1-n^2}$$

can be comparable to that observed for type III solar radio bursts, where δ reaches 0.1–0.2 (for example, $\delta \approx 0.2$ corresponds to $n \approx 0.1$).

Using limiting condition (7) with respect to the Cherenkov radiation frequency may probably provide some explanation for both the existence of radiation at a double frequency ($\omega \approx 2\omega_0$) and the observed misfit of the visible positions of radiation at the main ($\omega \approx \omega_0$) and doubled ($\omega \approx 2\omega_0$) frequency, since both ω_0 and n (determining δ) depend on the plasma density and, hence, on altitude. In the limit $n \longrightarrow 1$ corresponding to the outer layers of solar corona, relation (7) gives $\omega > \omega_0/2$, which may account for type V solar radio bursts observed in a broad frequency range, accompanying with a small delay (in ~10% of cases) type III bursts [27]. At present, type V bursts are explained (see, e.g., [27]) using an additional assumption that sufficiently strong magnetic fields exist in the corresponding spatial region, which are capable of inducing synchrotron radiation from relativistic particles. This radiation exhibits, similar to Cherenkov radiation, a characteristic anisotropy, but has a different nature like a kind of bremsstrahlung radiation.

It should be noted that the phenomenon of reverse frequency drift [27] and the corresponding increase in the radio wave frequency with time observed for type U solar radio bursts (also conventionally associated with strong magnetic fields) permit an explanation (even for retained direction of the electron beam) based on the unlimited expansion (following from inequalities (7)) of the permissible Cherenkov radiation frequency range for $n \longrightarrow 1$, whereby the beam moves toward the periphery of the solar corona.

The proposed theory of Cherenkov radiation, in contrast to the classical theory [5–11], can be used for independent consideration of many other problems in astrophysics and solar physics (e.g., the formation of energetics and structure of the solar chromosphere and corona). As has already been noted above, the proposed mechanism of Cherenkov radiation for $n < 1$ offers an important (especially in the region of low densities and weak fields) alternative to different variants of bremsstrahlung radiation usually considered for explanation of the observed generation of hard γ and RG radiation by high-energy charged particles of cosmic rays. Another interesting problem requiring separate consideration is elucidation of the physical mechanisms providing significant acceleration of such particles taking into account the limiting condition (16) (see also Appendix, Eq. (A.1), which admits (for $y < 0$) the possibility of such acceleration related to Cherenkov photon emission. Indeed, the existing theory of particle acceleration by radiation involves the possibility of acceleration, but only in the course of Cherenkov absorption of radiation [29, 36, 37]. For example, Tsytovich [29] employed formula (1) (or [11, Eq. (11)]) for determining the conditions under which Cherenkov absorption is possible (i.e., $\varepsilon_p < 0$ and $|\cos\theta| \le 1$ in (1)), while Cherenkov radiation is prohibited (for $\varepsilon_p > 0$ and $|\cos\theta| > 1$ in (1)).[2] Such interpretations do not take into account the aforementioned possibility of particle acceleration with simultaneous energetically favorable nonbremsstrahlung emission of transverse Cherenkov photons by the medium.

ACKNOWLEDGMENTS

The author is grateful to B.M. Bolotovskiĭ for fruitful discussions and to the referee for constructive questions.

This study was supported by the Russian Foundation for Basic Research, project nos. 01-05-64300 and 01-07-90211.

APPENDIX

1. For $y < 0$ (i.e., for $y = -|y|$) at $n < 1$ and $|y| < n$, condition (6) yields the following limitation with respect to v_0 that is necessary for Cherenkov photon emission with simultaneous acceleration of the electron:

$$\frac{n|y| - A}{B} < \beta < \frac{|y|}{n} \text{ if } \varepsilon < \frac{2|y|}{n^2 - y^2},$$
$$\frac{A - n|y|}{B} < \beta < \frac{|y|}{n} \text{ if } \frac{2|y|}{n^2 - y^2} < \varepsilon < \frac{4n|y|}{(n^2 - y^2)^{3/2}}, \quad \text{(A.1)}$$

where

$$\beta = \frac{v_0}{c}, \quad A = \frac{\varepsilon}{2}(n^2 - y^2)^{3/2}\left(1 + \frac{\varepsilon^2}{4}(n^2 - y^2)\right)^{1/2},$$

$$B = n^2 + \frac{\varepsilon^2}{4}(n^2 - y^2)^2.$$

Relations (A.1) admit realization of even hard Cherenkov radiation with $\varepsilon \gg 1$ for $|y| \longrightarrow n$ (that is, in the limit of $\tilde{\omega}_0/\omega \longrightarrow 1 + n$).

The obtained relativistic generalization of the Landau criterion in the form of conditions (16) for $y < 0$ leads to the possibility of energetically favorable

[2] There is either a misprint or error in [29, Eq. (15)]: the denominator contains $\delta\omega^{(0)}$ instead of $\delta\omega$, which affects the result.

realization of anisotropic Cherenkov radiation, replacing (A.1) by the following conditions:

$$\beta > \frac{|y|}{n} \text{ for } \varepsilon < \frac{4ny}{(n^2 - y^2)^{3/2}},$$

$$\beta > \frac{A - n|y|}{B} \text{ for } \varepsilon > \frac{4ny}{(n^2 - y^2)^{3/2}}.$$

However, condition (16) is satisfied for any $\cos\theta \geq -1$ (i.e., when this condition does not determine any anisotropy of radiation) even for $|y| > \beta n$, which corresponds to (A.1) and obeys the necessary condition (6) for the realization of Cherenkov radiation.

2. For $y > 0$, inequality (16) corresponding to the obtained relativistic generalization of the Landay criterion may take place only when $y\varepsilon/z < 1$ at an above-threshold velocity $v_0 > v_e = cy\varepsilon/z$, where $z = \varepsilon/n$ for $n > 1$ and $z = \varepsilon n$ for $n < 1$. Let us compare this limitation for v_0 to the conditions with respect to this velocity following from inequality (10). For $z > \beta/\gamma_0$, inequality (10) is always satisfied provided positive definiteness of the radicand. In this case, we obtain the following limitations:

$$v_2 < v_0 < v_1 \text{ at } \varepsilon + 1 - \sqrt{1 + z^2} < \tilde{\varepsilon} < \varepsilon,$$

where

$$\tilde{\varepsilon} = \frac{\varepsilon_0}{m_e c^2}, \quad v_1 = \frac{cz}{\sqrt{1 + z^2}}, \quad v_2 = \frac{c\sqrt{\varepsilon y(\varepsilon y + 2)}}{1 + \varepsilon y}.$$

For $v_0 > v_1$, inequality (10) can in principle be satisfied for both $z > \varepsilon y$ and $z < \varepsilon y$. However, the condition $v_0 > v_e$ separates only the variant of $z > \varepsilon y$, which corresponds to the energetically favorable realization of Cherenkov radiation under conditions (10). In this case, inequality (10) gives a threshold limit for the particle velocity v_0 necessary for Cherenkov photon emission, which coincides with formula (8) and appears as $v_0 > v_3 = \beta_1 c$, where $v_3 > v_e$(for $n > 1$, substitute $1/n$ for n in the expression for β_1). In the case of $v_3 > v_1$, there are two separate intervals of v_0 ensuring energetically favorable realization of Cherenkov radiation: $v_2 < v_0 < v_1$ and $v_0 > v_3$, since $v_0 < v_1$ always implies that $v_e < v_2$. At the same time, for $v_3 < v_1$, the only condition for the energetically favorable realization of Cherenkov radiation following from (10) is $v_0 > v_2$, because in this case the threshold velocity obeys the condition $v_e < v_2$. In particular, for $m_p = \Delta M = \varepsilon_0/c^2$, according to formula (4), inequality (10) gives for $n \ll 1$ and $\tilde{\omega}_0 = \omega_0$ the following conditions for the energetically favorable realization of Cherenkov radiation:

$$v_0 > cn\tilde{\varepsilon} \text{ for } \tilde{\varepsilon} > 1,$$

$$v_0 > \frac{cn(1 + \tilde{\varepsilon})}{2} \text{ for } \tilde{\varepsilon} < 1.$$

REFERENCES

1. O. Heaviside, *Electrical Papers* (Macmillan, London, 1892), Vol. 2, p. 492.
2. B. M. Bolotovskiĭ, *Oliver Heaviside 1850–1925* (Nauka, Moscow, 1985), p. 152 [in Russian].
3. P. A. Cherenkov, Dokl. Akad. Nauk SSSR **2**, 451 (1934); Usp. Fiz. Nauk **68**, 377 (1959) [Sov. Phys. Usp. **2**, 195 (1959)].
4. S. I. Vavilov, Dokl. Akad. Nauk SSSR **2**, 457 (1934).
5. I. E. Tamm and I. M. Frank, Dokl. Akad. Nauk SSSR **14**, 107 (1937).
6. I. E. Tamm, J. Phys. (Moscow) **1**, 439 (1939); in *Collected Scientific Works* (Nauka, Moscow, 1975), Vol. 1, p. 79 [in Russian].
7. I. M. Frank, Dokl. Akad. Nauk SSSR **42**, 354 (1944).
8. I. E. Tamm, Usp. Fiz. Nauk **68**, 387 (1959).
9. I. M. Frank, Usp. Fiz. Nauk **68**, 397 (1959).
10. V. L. Ginzburg, *Theoretical Physics and Astrophysics*, 2nd ed. (Nauka, Moscow, 1981; Pergamon Press, Oxford, 1979).
11. V. L. Ginzburg, Usp. Fiz. Nauk **166**, 1033 (1996) [Phys. Usp. **39**, 973 (1996)].
12. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media*, 2nd ed. (Nauka, Moscow, 1982, p. 553; Pergamon Press, Oxford, 1984).
13. V. P. Zrelov, Zh. Éksp. Teor. Fiz. **45**, 291 (1963) [Sov. Phys. JETP **18**, 203 (1964)].
14. V. G. Levich, Yu. A. Vdovin, and V. A. Myamlin, *Course of Theoretical Physics* (Nauka, Moscow, 1971), Vol. 2, p. 436 [in Russian].
15. V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics*, 3rd ed. (Nauka, Moscow, 1989, pp. 327, 504; Pergamon Press, Oxford, 1982).
16. W. Pauli, *The Theory of Relativity* (Pergamon Press, Oxford, 1958; Mir, Moscow, 1983, p. 336).
17. I. M. Dremin, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 185 (2002) [JETP Lett. **76**, 151 (2002)].
18. D. V. Skobel'tsin, Usp. Fiz. Nauk **122**, 295 (1977) [Sov. Phys. Usp. **20**, 528 (1977)].
19. V. L. Ginzburg, Usp. Fiz. Nauk **122**, 325 (1977) [Sov. Phys. Usp. **20**, 546 (1977)].
20. L. D. Landau, Zh. Éksp. Teor. Fiz. **11**, 592 (1941).
21. S. Chandrasekhar, *Ellipsoidal Figures of Equilibrium* (Yale Univ. Press, New Haven, Conn., 1969; Mir, Moscow, 1973).

22. M. V. Nezlin, Usp. Fiz. Nauk **120**, 481 (1976) [Sov. Phys. Usp. **19**, 946 (1976)].

23. B. M. Bolotovskiĭ and S. N. Stolyarov, in *Einstein Collection* (Nauka, Moscow, 1980), pp. 73–130 [in Russian].

24. S. G. Chefranov, Pis'ma Zh. Éksp. Teor. Fiz. **73**, 312 (2001) [JETP Lett. **73**, 274 (2001)].

25. M. I. Ryazanov, Zh. Éksp. Teor. Fiz. **45**, 333 (1963) [Sov. Phys. JETP **18**, 232 (1964)].

26. A. Gaĭlitis and V. N. Tsytovich, Zh. Éksp. Teor. Fiz. **46**, 1726 (1964) [Sov. Phys. JETP **19**, 1165 (1964)].

27. S. A. Kaplan, S. B. Pikel'ner, and V. N. Tsytovich, *Physics of the Solar Atmospheric Plasma. Modern Problems in Physics* (Nauka, Moscow, 1977), p. 256 [in Russian].

28. Yu. S. Vladimirov, Zh. Éksp. Teor. Fiz. **45**, 251 (1963) [Sov. Phys. JETP **18**, 176 (1964)].

29. V. N. Tsytovich, Dokl. Akad. Nauk SSSR **142**, 319 (1962) [Sov. Phys. Dokl. **7**, 43 (1962)].

30. V. N. Tsytovich, Astron. Zh. **41**, 7 (1964) [Sov. Astron. **8**, 4 (1964)].

31. V. L. Ginzburg and S. N. Syrovatskiĭ, *The Origin of Cosmic Rays* (Akad. Nauk SSSR, Moscow, 1963; Gordon and Breach, New York, 1969).

32. A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskiĭ, *Methods of Quantum Field Theory in Statistical Physics*, 2nd ed. (Fizmatgiz, Moscow, 1998, p. 26; Prentice Hall, Englewood Cliffs, N.J., 1963).

33. B. E. Nemtsov, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **28**, 1549 (1985).

34. E. B. Sonin, Usp. Fiz. Nauk **137**, 267 (1982) [Sov. Phys. Usp. **25**, 409 (1982)].

35. R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics* (Addison-Wesley, Reading, Mass., 1968; Mir, Moscow, 1977, pp. 233–243, 307–311), Vol. 6.

36. V. N. Tsytovich, *Lectures on Nonlinear Plasma Kinetics* (Springer, Berlin, 1995).

37. *Encyclopedia of Low-Temperature Plasma*, Ed. by V. E. Fortov (Nauka, Moscow 2000) [in Russian].

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 2, 2004, pp. 311–319.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 2, 2004, pp. 349–358.

ATOMS, SPECTRA, RADIATION

Diffraction Radiation from an Inhomogeneous Dielectric Film on the Surface of a Perfect Conductor

M. I. Ryazanov, M. N. Strikhanov, and A. A. Tishchenko*

Moscow Institute of Engineering Physics (State University), Kashirskoe sh. 31, Moscow, 115409 Russia

**e-mail: altis2001@mail333.com*

Received December 10, 2003

Abstract—Diffraction radiation generated by a charged particle moving uniformly parallel to the surface of a perfect conductor coated with a dielectric film is considered; the thickness of the film is an arbitrary function of coordinates. A particular case is considered when this function is periodic in one coordinate. The dependence of radiation on the profile of an individual irregularity of the periodic film is analyzed in detail for an arbitrary energy of the particle. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is well known that a charged particle moving uniformly in vacuum parallel to the plane surface of a homogeneous substance emits radiation only when the Vavilov–Cherenkov radiation conditions are fulfilled. If these conditions are not fulfilled, then there is no radiation because the transfer of the longitudinal (along the velocity vector) momentum is impossible. In the case of an inhomogeneous surface, the transfer of longitudinal momentum to the irregularities of the substance becomes possible and radiation is emitted, which is usually called diffraction radiation [1]. Various special cases of such radiation have been considered by many authors [2–6]; however, diffraction radiation that arises when a charged particle moves uniformly in vacuum parallel to the surface of a homogeneous semi-infinite conductor coated with an inhomogeneous dielectric film has not been considered. The aim of the present paper is to consider such radiation in the case when there is no Vavilov–Cherenkov radiation.

This kind of radiation is associated with an inhomogeneous film because radiation is not emitted in the absence of film. To satisfy the boundary conditions on the surface of the conductor, one introduces an image charge [7, 8]. The field outside the conductor coincides with the total field induced by two charges that move uniformly in vacuum: a real charge e and a fictitious charge $-e$. This field polarizes the inhomogeneous film, and the polarization currents induced in the film give rise to diffraction radiation. It is of interest to evaluate the intensity of the diffraction radiation as a function of the character of irregularities in the film.

2. POLARIZATION CURRENT IN THE SURFACE AREA OF A DIELECTRIC FILM

Denote by $\mathbf{E}_0(\mathbf{r}, t)$ the electric field induced in the half-space $x > 0$ by a uniformly moving particle with charge e. The charged particle moves parallel to the surface $x = 0$ of a perfect conductor according to the law $\mathbf{r} = \mathbf{a} + \mathbf{v}t$ (the vector $\mathbf{a}$ is perpendicular to the surface of the conductor). This field induces polarization currents in the inhomogeneous film, which give rise to diffraction radiation. If we neglect the interaction between the currents induced in the inhomogeneous film by the field of the traveling charged particle, we can assume that the field acting on a separate molecule in the film is equal to $\mathbf{E}_0(\mathbf{r}, t)$. Then, in the dipole approximation, the Fourier image of the polarization current is given by

$$\mathbf{j}(\mathbf{r}, \omega) = -i\omega n_{\text{mic}}(\mathbf{r})\alpha(\omega)\mathbf{E}_0(\mathbf{r}, \omega), \tag{1}$$

where $\alpha(\omega)$ is the polarizability of an individual molecule and $n_{\text{mic}}(\mathbf{r})$ is the microscopic density of molecules in the film.

The field $\mathbf{E}_0$ is determined by the method of images as a sum of the field of the real charge e and the field of the fictitious image-charge $-e$, which moves according to the law $\mathbf{r} = -\mathbf{a} + \mathbf{v}t$. Choosing axis z along $\mathbf{v}$ and axis x along $\mathbf{a}$, we can represent the Fourier image of the total field of the real charge and the image charge as

$$\mathbf{E}_0(\mathbf{r}, \omega) = i\int d^3q \exp\{i\mathbf{q}\cdot\mathbf{r}\} \times \mathbf{E}_0(q_x, q_y)\delta(\omega - q_z v)\sin(q_x a), \tag{2}$$

$$\mathbf{E}_0(q_x, q_y) = \frac{ie}{\pi^2}\frac{\mathbf{Q} + \mathbf{v}\dfrac{\omega}{\nu^2\gamma^2}}{Q^2 + \left(\dfrac{\omega}{\nu\gamma}\right)^2}, \qquad (3)$$

$$\frac{1}{\gamma^2} = 1 - \frac{\nu^2}{c^2}, \quad \mathbf{Q} = (q_x, q_y).$$

The substitution of (2) into (1) yields

$$\mathbf{j}_0(\mathbf{k}, \omega) = \frac{\omega\alpha(\omega)}{(2\pi)^3}\int d^3q\mathbf{E}_0(q_x, q_y)\delta(\omega - q_z\nu) \times \sin(q_x a)\sum_a \exp\{i(\mathbf{q} - \mathbf{k})\cdot\mathbf{R}_a\}, \qquad (4)$$

where the summation is performed over all molecules of the film. The field $\mathbf{E}_0$ does not contain the radiation field, so that the diffraction radiation is generated by the polarization current (4) in the inhomogeneous film. The angular and frequency distribution of the energy emitted by a given current in vacuum is given by

$$\frac{d^2E(\mathbf{n}, \omega)}{d\omega d\Omega} = (2\pi)^6\frac{1}{c}|\mathbf{k}\times\mathbf{j}_0(\mathbf{k}, \omega)|^2, \qquad (5)$$

where $\mathbf{k} = \mathbf{n}\omega/c$. This distribution is valid for the waves that reach a detector without reflection from the surface of the medium. However, one should also take into consideration the waves that reach the detector after reflection from the surface of the medium.

It is obvious that reflected waves are equivalent to the waves emitted by the polarization current of image molecules. Just as radiation in a real molecule is emitted by bound electrons accelerated by an external field, radiation in an image molecule is emitted by the images of bound electrons. The Fourier image of the current due to image molecules can be determined from (4) by the substitutions $X_a \longrightarrow -X_a$, $j_x \longrightarrow j_x$, $j_y \longrightarrow -j_y$, and $j_z \longrightarrow -j_z$. The first substitution yields a mirror-symmetric position of an image molecule with respect to the surface of the conductor. The remaining substitutions are associated with the application of the method of images to the electrons bound in a molecule. They correspond to a change in the sign of the velocity component normal to the surface and the sign of the charges bound in the image molecule. Hence, we obtain

$$\mathbf{j}_1(\mathbf{k}, \omega) = \frac{\omega\alpha(\omega)}{(2\pi)^3} \times \int d^3q\mathbf{E}_0(q_x, q_y)\delta(\omega - q_z\nu)\sin(q_x a) \times \sum_a \exp\{i(\mathbf{q} - \mathbf{k})\cdot\mathbf{R}_a + 2i(\mathbf{e}\cdot\mathbf{k})(\mathbf{e}\cdot\mathbf{R}_a)\}, \qquad (6)$$

where $\mathbf{e}$ is a unit vector normal to the surface, i.e., $\mathbf{e} \equiv \mathbf{e}_x$. The total energy in frequency interval $d\omega$ emitted into an tooth $d\Omega$ of solid angle, including the radiant energy reflected from the conductor surface, coincides with the distribution of radiant energy emitted in vacuum by two sources (4) and (6):

$$\frac{d^2E(\mathbf{n}, \omega)}{d\omega d\Omega} = (2\pi)^6\frac{1}{c}|\mathbf{k}\times(\mathbf{j}_0(\mathbf{k}, \omega) + \mathbf{j}_1(\mathbf{k}, \omega))|^2. \qquad (7)$$

3. DIFFRACTION RADIATION FROM SEPARATE ADSORBED ATOMS

Consider a limiting case of a film consisting of two adatoms situated at the points $\mathbf{R}_a = (d, 0, 0)$ and $\mathbf{R}_b = (d, 0, L)$ at a distance of L from each other. Then, the Fourier image of the density of the total polarization current is given by a sum of two terms related to the first and the second adatoms, and the spectral–angular distribution of the emitted energy has the form

$$\frac{d^2E(\mathbf{n}, \omega)}{d\omega d\Omega} = \frac{4\omega^2|\alpha(\omega)|^2}{c\nu^2} \times 2\left\{1 + \cos\left[L\left(\frac{\omega}{\nu} - k_z\right)\right]\right\}\cos^2(k_x d)|\mathbf{k}\times\mathbf{I}|^2, \qquad (8)$$

where

$$\mathbf{I} = \int_{-\infty}^{\infty} dq_x \int_{-\infty}^{\infty} dq_y \mathbf{E}_0(q_x, q_y)\sin(aq_x)\exp\{idq_x\}. \qquad (9)$$

One can see that the difference between the energies radiated into the interval $d\omega d\Omega$ for one and two adatoms consists in the appearance of the additional factor $2\{1 + \cos[L(\omega/\nu - k_z)]\}$ in the case of two adatoms.

When $|\omega/\nu - k_z|L \ll 1$, the energy arriving at a given observation point from two adatoms is greater by a factor of four than the energy from a single adatom. This is associated with the fact that the waves emitted by each adatom arrive at the observation point with equal phases, so that the field is doubled and the energy is quadrupled. Note that, when $L = 2sl$, where $l = \pi/|\omega/\nu - k_z|$ and s is an integer, the energy of diffraction radiation emitted by two adatoms is greater by a factor of four than the energy emitted by a single adatom. However, when $L = (2s + 1)l$, the energy of diffraction radiation emitted in the same direction is equal to zero. This is associated with the fact that the waves emitted by different adatoms arrive at the observation point with opposite phases and cancel out. Hence, if N adatoms lie on a line parallel to the velocity of the particle within the interval of length $l \sim \pi/|\omega/\nu - k_z|$, then the waves emitted by each adatom arrive at the observation point with equal phases, whereby the radiated energy proves to be greater by a factor of N^2 than that in the case of a single adatom. In the limit of $|\omega/\nu - k_z|L \gg 1$, the energy

emitted by two adatoms is greater by a factor of two than that emitted by a single adatom because the function $\cos[L(\omega/v - k_z)]$ rapidly oscillates in a frequency interval and gives a negligible contribution to the total radiation. The length l of a segment that corresponds to the coherent radiation of all adatoms situated in this segment is usually called coherence length. In the case of ultrarelativistic particles, whose energy $E \equiv \gamma mc^2 \gg mc^2$, the radiation is concentrated in the region of small angles θ between **n** and **v**, so that the coherence length is

$$l = \frac{\pi}{|\omega/v - k_z|} \sim \frac{2\pi c}{\omega(\theta^2 + \gamma^{-2})} \tag{10}$$

and may reach macroscopic values.

To determine an explicit expression for the frequency and angular distribution of diffraction radiation from two adatoms, we calculate integral (9). This integral can be expressed in terms of modified Bessel functions of the zeroth (K_0) and first (K_1) orders:

$$\mathbf{I} = (I_x, 0, I_z),$$

$$I_x = i\frac{e}{\pi}\frac{\omega}{v\gamma}$$

$$\times\left[K_1\left(\frac{\omega}{v\gamma}(a-d)\right) + K_1\left(\frac{\omega}{v\gamma}(a+d)\right)\right], \tag{11}$$

$$I_z = \frac{e}{\pi}\frac{\omega}{v\gamma^2}$$

$$\times\left[K_0\left(\frac{\omega}{v\gamma}(a-d)\right) - K_0\left(\frac{\omega}{v\gamma}(a+d)\right)\right].$$

In the case of an ultrarelativistic particle, the condition $I_x \gg I_z$ is fulfilled. Therefore,

$$|\mathbf{k}\times\mathbf{I}|^2 = k^2 I_x^2(1-n_x^2)$$

and the angular and frequency distribution of radiated energy for $\gamma \gg 1$ is given by

$$\frac{d^2E(\mathbf{n},\omega)}{d\omega d\Omega} = \frac{8\omega^4|\alpha(\omega)|^2}{c\beta^2}$$

$$\times\left\{1 + \cos\left[L\left(\frac{\omega}{v} - k_z\right)\right]\right\}\cos^2(k_x d) \tag{12}$$

$$\times(1-n_x^2)\frac{1}{h^2}\left\{\frac{\omega h}{v\gamma}K_1\left(\frac{\omega h}{v\gamma}\right) + \frac{\omega h}{v\gamma}K_1\left(\frac{\omega}{v\gamma}(h+2d)\right)\right\}^2.$$

Here, $h \equiv a - d$ is the impact parameter, i.e., the shortest distance between the particle trajectory and the molecules. Let us point out certain features of the distribution obtained.

First, the expression in braces is on the order of unity when $\omega h < v\gamma$ and rapidly vanishes for $\omega h > v\gamma$. Therefore, the radiated energy is inversely proportional to the square of the impact parameter h for $h < v\gamma/\omega$ and exponentially tends to zero for $h > v\gamma/\omega$.

Second, the forward radiation ($n_z = 1$, $n_x = 0$) does not vanish. This is associated with the fact that the component of the polarization current normal to the surface, $j_x \propto I_x$, is different from zero.

Third, the amount of energy radiated at a certain fixed angle at a fixed frequency oscillates as a function of d, which is the distance from the molecules to the surface of the conductor. These oscillations result from the interference between the radiation fields of the real and image molecules.

4. DIFFRACTION RADIATION FROM THE SURFACE LAYER

Now, consider the diffraction radiation from an inhomogeneous film on the surface of a semi-infinite conductor. Suppose that the mean density $n(\mathbf{r})$ of conduction electrons in the film as a function of coordinates is known. In macroscopic electrodynamics, the Fourier images of currents $\mathbf{j}_0(\mathbf{k}, \omega)$ and $\mathbf{j}_1(\mathbf{k}, \omega)$ in (7) can be replaced by their mean values $\langle\mathbf{j}_0(\mathbf{k}, \omega)\rangle$ and $\langle\mathbf{j}_1(\mathbf{k}, \omega)\rangle$:

$$\langle\mathbf{j}_0(\mathbf{k},\omega)\rangle = \alpha(\omega)\int d^3q\mathbf{E}_0(q_x, q_y)$$

$$\times\delta(\omega - q_z v)\sin(aq_x)n(\mathbf{k}-\mathbf{q}),$$

$$\langle\mathbf{j}_1(\mathbf{k},\omega)\rangle = \alpha(\omega)\int d^3q\mathbf{E}_0(q_x, q_y) \tag{13}$$

$$\times\delta(\omega - q_z v)\sin(aq_x)n(\mathbf{k}-\mathbf{q}-2\mathbf{e}(\mathbf{e}\cdot\mathbf{k})).$$

Here,

$$n(\mathbf{k}-\mathbf{q}) = \int\frac{d^3r}{(2\pi)^3}n(\mathbf{r})\exp\{-i\mathbf{r}\cdot(\mathbf{k}-\mathbf{q})\}$$

is the Fourier image of the density of molecules. Thus, the angular and frequency distribution of radiated energy can be expressed as

$$\frac{d^2E(\mathbf{n},\omega)}{d\omega d\Omega} = \frac{\omega^2}{c}|\alpha(\omega)|^2|\mathbf{k}\times\mathbf{J}(\mathbf{k},\omega)|^2, \tag{14}$$

where $\mathbf{J}(\mathbf{k}, \omega)$ is defined as

$$\mathbf{J}(\mathbf{k},\omega) = (2\pi)^3\int d^3q\mathbf{E}_0(q_x, q_y)\sin(aq_x)$$
$$\times\delta(\omega - q_z v)\{n(\mathbf{k}-\mathbf{q}) + n(\mathbf{k}-\mathbf{q}-2\mathbf{e}(\mathbf{e}\cdot\mathbf{k}))\}. \tag{15}$$

The radiation distribution (14) is valid for an arbitrary dependence of the density of molecules in the film on coordinates.

Let us consider, as an example, a film that is inhomogeneous along one coordinate. Suppose that the film profile is given by a function $g(z)$. The mean density of molecules is obtained by averaging the microscopic density over the coordinates of all molecules in the film (n is the mean density of molecules in the film):

$$n(\mathbf{r}) = \langle n^{\mathrm{mic}}(\mathbf{r})\rangle = \left\langle \sum_b \delta(\mathbf{r}-\mathbf{R}_b)\right\rangle = n\int_{-\infty}^{\infty} dY_b \int_{-\infty}^{\infty} dZ_b \int_0^{g(Z_b)} dX_b \delta(\mathbf{r}-\mathbf{R}_b). \tag{16}$$

Taking the Fourier image of (16) and substituting the result into (15), one can obtain

$$\mathbf{J}(\mathbf{k},\omega) = \frac{4en}{\nu}\frac{\exp\{-a\eta\}}{\eta^2+k_x^2} \times\left(iG_1\mathbf{e}_x - \frac{G_2}{\eta}\left(k_y\mathbf{e}_y + \frac{\omega}{\nu\gamma^2}\mathbf{e}_z\right)\right), \tag{17}$$

where

$$\varphi \equiv \frac{\omega}{\nu} - k_z = k(\beta^{-1} - n_z), \tag{18}$$

$$\eta = \frac{\omega}{c\beta\gamma}\sqrt{1+\gamma^2\beta^2 n_y^2}, \tag{19}$$

$$G_1 = \int_{-\infty}^{\infty} dz \exp\{iz\varphi\}\{\eta\cos[k_x g(z)] \times \sinh[\eta g(z)] + k_x \sin[k_x g(z)]\cosh[\eta g(z)]\},$$

$$G_2 = \int_{-\infty}^{\infty} dz \exp\{iz\varphi\}\{\eta\cos[k_x g(z)] \times \cosh[\eta g(z)] - \eta + k_x \sin[k_x g(z)]\sinh[\eta g(z)]\}. \tag{20}$$

Here, G_1 and G_2 are factors defined by the surface profile $g(z)$ of the dielectric film.

Consider, for definiteness, the particular case of an ultrarelativistic charged particle $\gamma \gg 1$, moderately small radiation angles $n_x \gg \gamma^{-1}$, and a thin film of thickness $g_{\max}(z) = b \ll \gamma c/\omega$. Note that the radiation is mainly concentrated in the region $n_y \leq \gamma^{-1}$ on the plane $n_y = 0$ in view of the factor

$$\exp\left\{-\frac{a\omega}{c\beta\gamma}\sqrt{1+\gamma^2\beta^2 n_y^2}\right\}.$$

For simplicity, we further consider the radiation only in the plane $n_y = 0$. In this case, formula (17) is substantially simplified:

$$\mathbf{J}(\mathbf{k},\omega) = i\frac{4en}{c}\mathbf{e}_x \exp\{-a\eta\} \times \int_{-\infty}^{\infty} dz \exp\{iz\varphi\} g(z). \tag{21}$$

The substitution of (21) into (14) yields the spectral–angular distribution of energy radiated by an ultrarelativistic charged particle traveling over a thin inhomogeneous dielectric film placed on a perfectly conducting substrate:

$$\frac{d^2E(\mathbf{n},\omega)}{d\omega d\Omega} = \frac{e^2}{c}|4n\alpha(\omega)|^2(1-n_x^2)\exp\{-2a\eta\} \times\left|\frac{\omega^2}{c^2}\int_{-\infty}^{\infty} dz \exp\{iz\varphi\} g(z)\right|^2. \tag{22}$$

Now, suppose that the film is periodic. Then, the function $g(z)$ describing the profile of the film can be expanded into the Fourier series

$$g(z) = \sum_m g_m \exp\left\{iz\frac{2\pi}{p}m\right\}, \tag{23}$$

where p is the period of the film and the coefficients g_m are defined by the shape of a separate irregularity. In this case, the integral in (22) can be transformed into

$$\int_{-\infty}^{\infty} dz \exp\{iz\varphi\}\sum_m g_m \exp\left\{iz\frac{2\pi}{p}m\right\} = 2\pi\sum_m g_m \delta\left(\frac{\omega}{\nu} - k_z + \frac{2\pi}{p}m\right). \tag{24}$$

The sum of delta functions in this formula indicates that the spectrum contains sharp lines at the frequency points satisfying the relation

$$\frac{m\lambda}{p} = \beta^{-1} - n_z, \quad m = 1, 2, \ldots. \tag{25}$$

Physically, formula (25) has its roots in the fact that the traveling particle can transfer only discrete momentum to a periodically inhomogeneous medium. The diffraction radiation associated with the periodicity of a medium near which a fast charged particle passes by is for brevity called resonance diffraction radiation [6].

For $m = 1$, formula (25) implies the well-known Smith–Purcell relation. For a fixed period p and wavelength λ, the diffraction order is bounded by the two-sided inequality

$$\frac{p}{\lambda}(\beta^{-1} - 1) \leq m \leq \frac{p}{\lambda}(1 + \beta^{-1}). \tag{26}$$

For example, when $p = 0.833$ μm, $\lambda = 0.360$ μm, and $\gamma = 1673$, one obtains $m = 1, 2, 3$ and $\theta = 55.4°$, 82.2°, 107.3°, respectively. The Smith–Purcell radiation with these parameters has recently been observed by a group of German physicists with a MAMI microtron [9]. They used a periodically deformed surface (glass BK7) spray-coated with a thin (~0.7 μm) metal film (aluminum).

One can see that, in general, the angles at which the resonance radiation propagates are large. When $\gamma \gg 1$ and $p < \lambda/2$, radiation at angles defined by formulas (25) is suppressed because there is no m satisfying condition (26).

Inequalities (26) show that resonance radiation can be generated by charged particles of any energy. For example, diffraction radiation of nonrelativistic electrons is widely used in microwave electronics [10]. The spectrum of the resonance diffraction radiation of relativistic charged particles includes optical, millimeter-wave, and submillimeter-wave bands [11, 12].

5. RADIATION FROM A DIFFRACTION GRATING OF ARBITRARY PROFILE

Now, let us consider a particular case of a surface layer in the form of a finite diffraction grating on the surface of a perfect conductor occupying the half-space $x < 0$ (see figure). Suppose that the grating is uniform along the y axis and its profile in the z direction consists of N periodic teeth. The period of the grating is p, the width of an tooth is w, and its height is given by a function $f(z)$. The mean density of molecules is the same in all teeth and equal to n.

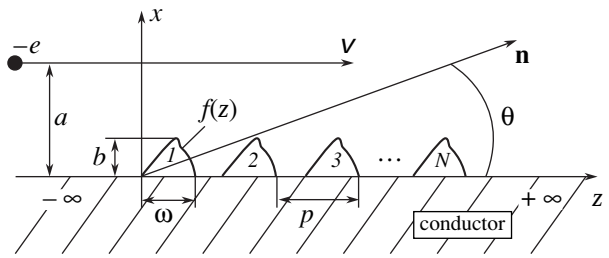
The mean density of molecules is determined by averaging the microscopic density over the coordinates of all molecules of the film:

$$n(\mathbf{r}) = \langle n^{\mathrm{mic}}(\mathbf{r})\rangle = \left\langle \sum_{s=0}^{N-1}\sum_{b}\delta(\mathbf{r}-\mathbf{R}_s-\mathbf{R}_{bs})\right\rangle$$
$$= n\sum_{s=0}^{N-1}\int_{-\infty}^{\infty} dY_{bs}\int_0^w dZ_{bs}\int_0^{f(Z_{bs})} dX_{bs}\,\delta(\mathbf{r}-\mathbf{R}_s-\mathbf{R}_{bs}).$$

The vector $\mathbf{R}_s = \{0, 0, sp\}$ in the argument of the delta function takes into account the periodicity of teeth; s is the number of an tooth. Hence,

$$n(\mathbf{r}) = n\sum_{s=0}^{N-1}\xi(x, z-sp),$$
$$\xi(x, z-sp) = \int_0^w dz'\,\delta(z-sp-z')\int_0^{f(z')} dx'(x-x'), \tag{27}$$

so that the Fourier image of the density of molecules in the film is given by



Geometry of the problem. Diffraction grating consists of N dielectric strips on a conducting substrate. The strips are parallel to the y axis. The grating has the form of N teeth in the cross section made by the plane xz. The profile of a separate tooth is described by the function $f(z)$. A charged particle travels along the z axis.

$$n(\mathbf{q}) = n\delta(q_y)\xi(q_x, q_z)\sum_{s=0}^{N-1}\exp\{-isp q_z\}. \tag{28}$$

The substitution of (28) into (15) and integration with respect to q_z and q_y yields

$$\mathbf{J}(\mathbf{k}, \omega) = \frac{n}{\nu}\sum_{s=0}^{N-1}\exp\{-isp\varphi\}\int_{-\infty}^{\infty} dq_x\mathbf{E}_0(q_x, k_y)$$
$$\times \sin(aq_x)[\xi(k_x-q_x, -\varphi) + \xi(-k_x-q_x, -\varphi)].$$

Introducing

$$F_1 = \int_0^w dz\exp\{iz\varphi\}\{\eta\cos[k_x f(z)]\sinh[\eta f(z)]$$
$$+ k_x\sin[k_x f(z)]\cosh[\eta f(z)]\},$$
$$F_2 = \int_0^w dz\exp\{iz\varphi\}\{\eta\cos[k_x f(z)]\cosh[\eta f(z)] \tag{29}$$
$$-\eta + k_x\sin[k_x f(z)]\sinh[\eta f(z)]\},$$

where φ and η are defined by formulas (18) and (19), respectively, we can reduce the expression for $\mathbf{J}$ to

$$\mathbf{J}(\mathbf{k}, \omega) = \frac{4en}{\nu}\frac{\exp\{-a\eta\}}{\eta^2 + k_x^2}$$
$$\times\left(iF_1\mathbf{e}_x - \frac{F_2}{\eta}\left(k_y\mathbf{e}_y + \frac{\omega}{\nu\gamma^2}\mathbf{e}_z\right)\right)\sum_{s=0}^{N-1}\exp\{isp\varphi\}. \tag{30}$$

Here, F_1 and F_2 are factors defined by the profile of an

individual tooth. The substitution of (30) into (14) yields

$$\frac{d^2E(\mathbf{n},\omega)}{d\omega d\Omega} = \frac{e^2}{c}\frac{16\beta^2 n^2|\alpha(\omega)|^2}{(1-\beta^2 n_z^2)^2}\frac{\sin^2(\varphi pN/2)}{\sin^2(\varphi p/2)}$$

$$\times\left|\mathbf{n}\times\left(iF_1\mathbf{e}_x - F_2\frac{\beta\gamma n_y\mathbf{e}_y + \gamma^{-1}\mathbf{e}_z}{\sqrt{1+\gamma^2\beta^2 n_y^2}}\right)\right|^2 \qquad (31)$$

$$\times\exp\left\{-\frac{2a\omega}{c\beta\gamma}\sqrt{1+\gamma^2\beta^2 n_y^2}\right\}.$$

This formula describes the spectral–angular density of radiated energy in the case when the surface of a perfect conductor is coated with a periodically inhomogeneous dielectric film. In this case, the particle energy may be arbitrary. The properties of the medium are characterized by (1) the polarizability $\alpha(\omega)$ of an individual molecule; (2) the parameters that define the geometry of a separate irregularity (an tooth): the width w and the profile $f(z)$ of the tooth; and (3) the parameters that define the structure of the whole grating: the grating period p and the number of teeth N.

The spectral–angular characteristics of radiation (31) are determined by two factors. The first factor is $(1-\beta^2 n_z^2)^{-2}$, which gives a maximal forward radiation at small angles on the order of γ^{-1}. The second factor is given by the function

$$\frac{\sin^2(\varphi pN/2)}{\sin^2(\varphi p/2)}.$$

This function attains its maxima at the points $\varphi p = 2\pi m$, $m = 1, 2, 3, \ldots$; this yields a condition that coincides with the condition of resonance radiation (25). Introducing a polar angle θ by the relation $n_z = \cos\theta$, we obtain

$$m\frac{\lambda}{p} = \beta^{-1} - \cos\theta, \quad m = 1, 2, 3, \ldots. \qquad (32)$$

The angular width of separate peaks is $\Delta\theta = \lambda/Np$. When $N \gg 1$, the peaks become so sharp that the ratio of the squares of sines can be replaced by a sum of delta functions:

$$\frac{\sin^2(\varphi pN/2)}{\sin^2(\varphi p/2)} \xrightarrow{N\gg 1} 2\pi N\sum_m \delta(\varphi p - 2\pi m), \qquad (33)$$

which corresponds to the transition to a semi-infinite periodic film.

Now, let us analyze the nonrelativistic and ultrarelativistic cases separately.

6. NONRELATIVISTIC CASE

Consider a nonrelativistic particle with $\beta \ll 1$. In this case, the energy of the particle must be much greater than the total radiation loss since, otherwise, the assumption made above that the particle velocity is constant is violated. Formula (19) shows that $\eta = \omega/c\beta \gg k_x$ in this case. Then, we have three parameters with the dimension of length: λ, $\beta\lambda$, and $b = f_{\max}(z)$. Let us write out possible inequalities:

$$\beta\lambda \ll \lambda \ll b \longleftrightarrow k_x f(z) \gg 1, \quad \eta f(z) \gg 1, \qquad (34a)$$

$$\beta\lambda \ll b \ll \lambda \longleftrightarrow k_x f(z) \ll 1, \quad \eta f(z) \gg 1, \qquad (34b)$$

$$b \ll \beta\lambda \ll \lambda \longleftrightarrow k_x f(z) \ll 1, \quad \eta f(z) \ll 1. \qquad (34c)$$

The corresponding expressions for F_1 ad F_2 in (29) are given by

$$F_1 \approx F_2$$

$$\approx \frac{\eta}{2}\int_0^w dz\exp\{iz\varphi + \eta f(z)\}\cos[k_x f(z)], \qquad (35a)$$

$$F_1 \approx F_2 \approx \frac{\eta}{2}\int_0^w dz\exp\{iz\varphi + \eta f(z)\}, \qquad (35b)$$

$$F_2 \ll F_1 \approx \eta^2\int_0^w dz\exp\{iz\varphi\} f(z). \qquad (35c)$$

In the last case (35c), the spectral–angular distribution of radiated energy is expressed as

$$\frac{d^2E(\mathbf{n},\omega)}{d\omega d\Omega} = \frac{e^2}{c}n^2|\alpha(\omega)|^2(1-n_x^2)$$

$$\times\left|\frac{4\omega^2}{\beta c^2}\int_0^w dz\exp\{iz\varphi\} f(z)\right|^2 \qquad (36)$$

$$\times\frac{\sin^2(\varphi pN/2)}{\sin^2(\varphi p/2)}\exp\left\{-\frac{2a\omega}{c\beta}\right\}.$$

Instead of separate cases (34a) or (34b), it is more convenient to analyze a more general inequality

$$\eta b \gg 1 \qquad (37)$$

without any assumptions about the magnitude of $k_x b$. Here, instead of (35a) and (35b), we have

$$F_1 \approx F_2 \approx F$$

$$= \frac{\eta}{2}\int_0^w dz\exp\{iz\varphi + \eta f(z)\}\cos[k_x f(z)]. \qquad (38)$$

Consider (38) in greater detail. For positive values of the argument, the integrand $\exp\{\eta f(z)\}$ increases with increasing $f(z)$; therefore, under condition (37), we have

$$|F| = \left|\frac{\eta}{2}\int_0^w dz \exp\{iz\varphi + \eta f(z)\}\cos[k_x f(z)]\right|$$

$$\leq \frac{\eta}{2}\exp\{\eta f_{\max}\}\left|\cos(k_x f_{\max})\int_0^w dz\exp\{iz\varphi\}\right| \quad (39)$$

$$= \frac{\eta}{\varphi}\exp\{\eta b\}\left|\cos(k_x b)\sin\frac{w\varphi}{2}\right|.$$

This formula shows that, for a given width w and height b of an tooth, the radiation is maximal for rectangular teeth, when $f(z) = b$. In this case, the radiation for (38) is described by

$$\frac{d^2E(\mathbf{n},\omega)}{d\omega d\Omega}$$

$$= \frac{e^2}{c}\left(\frac{4\beta\eta}{\varphi}\right)^2 n^2|\alpha(\omega)|^2(1+n_y^2)\cos^2(k_x b) \quad (40)$$

$$\times \sin^2\frac{w\varphi}{2}\frac{\sin^2(\varphi pN/2)}{\sin^2(\varphi p/2)}\exp\left\{-\frac{2h\omega}{c\beta}\right\},$$

where $h \equiv a - b$ is the shortest distance between the radiating substance and the charged particle traveling over the substance, the so-called impact parameter.

We can evaluate integral (39) under fairly general assumptions on the explicit form of $f(z)$. To this end, we apply the Laplace method. Suppose that the function $f(z)$ has one maximum at a certain internal point z_0 of the interval $(0, w)$, so that $f(z_0) = f_{\max} = b$. In view of (37) and the inequality $\eta \gg k$, the main contribution to the integral (38) is made by a narrow domain of z near the point z_0. The oscillating terms exhibit rather smooth behavior near the point $z = z_0$, so that they can be taken outside the sign of integral. Expanding $f(z)$ in a series and retaining only the term quadratic in z, we obtain

$$F \approx \frac{\eta}{2}\sqrt{\frac{2\pi}{\eta|f''(z_0)|}}\exp\{iz_0\varphi\}\exp\{\eta b\}\cos(k_x b). \quad (41)$$

The spectral–angular distribution of radiation is expressed as

$$\frac{d^2E(\mathbf{n},\omega)}{d\omega d\Omega} = \frac{e^2}{c}\frac{8\pi\omega\beta}{c|f''(z_0)|}n^2|\alpha(\omega)|^2(1+n_y^2)$$

$$\times \cos^2(k_x b)\frac{\sin^2(\varphi pN/2)}{\sin^2(\varphi p/2)}\exp\left\{-\frac{2h\omega}{c\beta}\right\}. \quad (42)$$

In the case (34a), we have to replace the square of the cosine, $\cos^2(k_x b)$, by its mean value, 1/2, and, in the case (34b), by unity. Formula (42) describes the diffraction radiation from a grating with sufficiently high teeth, so that the inequality $b \gg \beta\lambda$ (37) holds.

7. ULTRARELATIVISTIC CASE

Now, consider the case $\gamma \gg 1$. It has been pointed out above that the radiation is largely concentrated in the domain $n_y \leq \gamma^{-1}$ in the plane $n_y = 0$. For simplicity, we will take into account only the radiation in the plane $n_y = 0$.

Since $iF_1\mathbf{e}_x \gg F_2\gamma^{-1}\mathbf{e}_z$, the main role in the ultrarelativistic case is played by the component of induced currents that is normal to the surface. For $n_y = 0$ and $\gamma \gg 1$, from (31) we obtain

$$\frac{d^2E(\mathbf{n},\omega)}{d\omega d\Omega} = \frac{e^2}{c}16n^2|\alpha(\omega)|^2$$

$$\times \frac{n_z^2|F_1|^2}{(\gamma^{-2}+n_x^2)^2}\frac{\sin^2(\varphi pN/2)}{\sin^2(\varphi p/2)}\exp\left\{-\frac{2a\omega}{c\gamma}\right\}. \quad (43)$$

The factor F_1 is determined by the first formula in (29) and depends on the explicit form of the function $f(z)$, which defines the profile of an individual tooth. Suppose that the profile of an tooth is given by a rectangle of width w and height b. Then, $f(z) = b$, and one can readily calculate $|F_1|^2$. For $n_y = 0$ and $\gamma \gg 1$, we have

$$\frac{d^2E(\mathbf{n},\omega)}{d\omega d\Omega} = \frac{e^2}{c}\frac{16n^2|\alpha(\omega)|^2 n_z^2}{(\gamma^{-2}+n_x^2)^2(1-\beta n_z)^2}\sin^2\frac{\varphi w}{2}$$

$$\times \frac{\sin^2(\varphi pN/2)}{\sin^2(\varphi p/2)}\exp\left\{-\frac{2h\omega}{c\gamma}\right\}$$

$$\times\left\{\gamma^{-1}\left[1-\exp\left(-\frac{2b\omega}{c\gamma}\right)\right]\cos(bk_x)\right. \quad (44)$$

$$\left. + n_x\left[1+\exp\left(-\frac{2b\omega}{c\gamma}\right)\right]\sin(bk_x)\right\}^2.$$

Here, as before, $h \equiv a - b$ is the impact parameter and $\varphi = (\omega/c)(\beta^{-1} - n_z)$.

Further, we will restrict the analysis to the resonance case. As was pointed out above, the resonance radiation propagates at sufficiently large angles (at any rate, it is this case that is of experimental interest). This fact suggests that $k_x \sim k \gg \eta$. Then, (29) implies that the factor F_1 for teeth of arbitrary shape can be expressed as

$$F_1 \approx k_x\int_0^w dz\exp\{iz\varphi\}\sin[k_x f(z)]\cosh[\eta f(z)]. \quad (45)$$

For thin plates of thickness less than the wavelength, $b \ll \lambda$, the following inequality holds:

$$b \ll \lambda \ll \gamma\lambda \longleftrightarrow \eta f(z) \ll k_x f(z) \ll 1. \quad (46)$$

Hence, (45) can be rewritten as

$$F_1 \approx k_x^2 \int_0^w dz \exp\{iz\varphi\} f(z). \quad (47)$$

For plates of thickness greater than the wavelength, one can assume that the inequality $k_x f(z) \gg 1$ holds. In view of the condition $k_x \sim k \gg \eta$, one cannot directly apply the Laplace method because the integrand in (45) rapidly oscillates near the maximum of the function $f(z)$. Let us represent the sine as a difference of exponents and apply the saddle-point method, i.e., move the integration contour to the complex plane so that it passes through the point z_+ (or z_-), which is a saddle point for the function $iz\varphi + ik_x f(z)$ (or, respectively, for $iz\varphi - ik_x f(z)$). In this case, the hyperbolic cosine can be taken outside the sign of integral at the point z_+ (or z_-) owing to the inequality $k_x \gg \eta$. From (45), one can easily derive

$$F_1 \approx \sqrt{\frac{\pi}{2}}\frac{k_x}{2i}\left\{\frac{\cosh[\eta f(z_+)]}{\sqrt{|g_+|}}\exp\{i\xi_+\}\right.$$
$$\times \exp\{iz_+\varphi + ik_x f(z_+)\} - \frac{\cosh[\eta f(z_-)]}{\sqrt{|g_-|}} \quad (48)$$
$$\left.\times \exp\{i\xi_-\}\exp\{iz_-\varphi - ik_x f(z_-)\}\right\}.$$

Here, the saddle points z_+ and z_- are given by the equation

$$k_x f'(z_\pm) \pm \varphi = 0. \quad (49)$$

The parameters $g_\pm$ and $\xi_\pm$ can be found from the equations

$$\pm ik_x f''(z_\pm) = g_\pm \exp\{i\psi_\pm\},$$
$$\cos(2\xi_\pm + \psi_\pm) = -1. \quad (50)$$

When deriving (49), we assumed that the function $f(z)$ attains its maximum at an internal point of the interval $(0, w)$ and that $f''(z_\pm) \neq 0$.

Note that, since the saddle points $z_\pm$ differ from z_0—the maximum point of the function $f(z)$ (i.e., the point where $f'(z_0) = 0$)—the factor F_1, obtained by formula (49), depends on $f(z_-)$ and $f(z_+)$ rather than on $f(z_0) = f_{max} = b$. Physically, this means that the maximal contribution to the radiation is given by some other points rather than by the vertices of the teeth. This fact may reduce the radiation because any point of an tooth lies farther from the trajectory of a traveling particle than the vertex. This effect will manifest itself when the inequality $|z_\pm - z_0| \geq \gamma\lambda$ holds, where $\gamma\lambda$ is a characteristic scale at which the proper field of the traveling particle decreases.

As an illustration, consider the function

$$f(z) = \frac{4bz}{w^2}(w - z).$$

This function describes an tooth in the form of a parabola with the vertex at the point $(w/2, b)$. From (49), we obtain the saddle points

$$z_\pm = \frac{w}{2}\left(1 \pm \frac{w\varphi}{2bk_x}\right) \quad (51)$$

and the corresponding values of the function $f(z_\pm)$:

$$f(z_\pm) = f \equiv b\left[1 - \left(\frac{w\varphi}{2bk_x}\right)^2\right]. \quad (52)$$

From formulas (50), we obtain

$$g_\pm = g \equiv \frac{8bk_x}{w^2},$$
$$\xi_\pm = \frac{1}{2}\left[\pi(2n+1) \pm \frac{\pi}{2}\right], \quad n = 0, 1, 2, \ldots. \quad (53)$$

Substituting (51)–(53) into (48), we obtain

$$|F|_1 \approx \sqrt{\frac{\pi}{2g}}\cosh(\eta f)\left|k_x \sin\left(\frac{\varphi^2 w^2}{4bk_x} - \frac{\pi}{4}\right)\right|. \quad (54)$$

Note that the applicability of (48) is restricted by the inequalities $bk_x \gg 1$ and $k_x \gg \eta \approx (\gamma\lambda)^{-1}$. It follows from (43), (52), and (54) that the radiation depends on

$$\exp\left\{-\frac{2\omega}{c\gamma}\left(h + b\left(\frac{w\varphi}{2bk_x}\right)^2\right)\right\}$$

rather than on $\exp\{-2h\omega/c\gamma\}$ as in the case of rectangular teeth (see (44)). However, for the resonance radiation, $\varphi p = 2\pi m$, $m = 1, 2, \ldots$, and

$$\frac{w\varphi}{2bk_x} = \frac{w\pi m}{pbk_x} \sim \frac{\pi}{bk_x} \ll 1$$

by virtue of the condition $bk_x \gg 1$. Therefore, in the case considered, the points z_+ and z_- are very close to z_0, and there is no appreciable attenuation of radiation.

Consider an explicit form of the coefficient F_1 corresponding to triangular teeth for $f(z) = zb/w$. Neglect-

ing the terms $\eta = \omega/c\gamma \ll k_x$ and calculating the integral in (45), we obtain the following expression for $|F_1|^2$:

$$|F_1|^2 = w^2 k_x^2 \cosh^2\left(\frac{\omega b}{c\gamma}\right)\left\{\frac{1}{a_1^2}\sin^2\frac{a_1}{2} + \frac{1}{a_2^2}\sin^2\frac{a_2}{2} - \frac{1 + \cos(a_1 - a_2) - \cos a_1 - \cos a_2}{2a_1 a_2}\right\}, \quad (55)$$

where

$$a_1 = \frac{\omega}{c}[w(1 - \beta n_z) + b n_x],$$

$$a_2 = \frac{\omega}{c}[w(1 - \beta n_z) + b n_x].$$

Substituting (55) into (43), we can obtain a spectral–angular distribution for the diffraction radiation of an ultrarelativistic particle in the case of triangular teeth.

8. DISCUSSION OF THE RESULTS

In this paper, we have investigated the diffraction radiation from an inhomogeneous dielectric film on the surface of a perfectly conducting substrate. To take into account the effect of the substrate, we have applied the method of images. Formulas (29) and (31) yield the spectral–angular density of radiation for a diffraction grating with an arbitrary profile $f(z)$ and a uniformly moving charged particle.

For a grating in which the height of an individual tooth b is less than any other parameter of the problem, the spectral–angular distribution of radiated energy as a function of the shape of a separate irregularity—an tooth—is determined by the factor

$$\int_0^w dz \exp\{iz\varphi\} f(z)$$

both in the nonrelativistic ($b \ll \beta\lambda$) and ultrarelativistic ($b \ll \lambda$) cases (see (47) and (36)).

For sufficiently thick plates (with a thickness of $b \gg \beta\lambda$), maximal radiation in the nonrelativistic case is attained for rectangular teeth. In the ultrarelativistic case, the dependence of $f(z)$ on the shape of a separate irregularity is more complicated, and it is difficult to make an unambiguous conclusion about a preferred shape of teeth. However, for rectangular teeth, the radiation distribution (31) is calculated most easily. For example, the radiation from rectangular plates is described by formulas (44) and (40) for the ultrarelativistic and nonrelativistic radiation cases, respectively.

These results have been obtained in the first order of perturbation theory by replacing the effective field by the sum of the fields of a traveling particle and its image. Second-order corrections to perturbation theory yield the following constraints on the applicability domain of the results obtained:

$$|\varepsilon - 1| \ll 1, \quad b \gg \gamma\beta\lambda,$$
$$|\varepsilon - 1|\frac{b}{\gamma\beta\lambda} \ll 1, \quad b \ll \gamma\beta\lambda. \quad (56)$$

Thus, in the transparency domain of the dielectric, where $|\varepsilon - 1| \ll 1$, the approximation used describes the radiation to a good accuracy for any thickness of the film. For a sufficiently thin film, when $b \ll \gamma\beta\lambda$, the applicability domain increases as the film thickness b decreases. The method applied has allowed us for the first time to take into account the effect of a substrate and analyze the dependence of diffraction radiation on the shape of separate irregularities (the grating).

The macroscopic averaging when deriving (14) is performed under the assumption that the number of molecules in a separate irregularity of the film is much greater than unity. This fact restricts from below the size of an individual irregularity. The applicability domain of the results obtained is also restricted by the assumption that the metal substrate is perfectly conducting. This fact restricts the range of frequencies from above.

REFERENCES

1. B. M. Bolotovskiĭ and G. V. Voskresenskiĭ, Usp. Fiz. Nauk **94**, 377 (1968) [Sov. Phys. Usp. **11**, 143 (1968)].
2. Yu. N. Dnestrovskiĭ and D. P. Kostomarov, Dokl. Akad. Nauk SSSR **116**, 377 (1957) [Sov. Phys. Dokl. **2**, 442 (1957)]; Dokl. Akad. Nauk SSSR **124**, 1026 (1959) [Sov. Phys. Dokl. **4**, 158 (1959)].
3. A. P. Kazantsev and G. I. Surdutovich, Dokl. Akad. Nauk SSSR **147**, 74 (1962) [Sov. Phys. Dokl. **7**, 990 (1963)].
4. J. H. Brownell, G. Doucas, and J. Walsh, Phys. Rev. E **57**, 1075 (1998).
5. A. P. Potylitsyn, Nucl. Instrum. Methods Phys. Res. B **145**, 169 (1998).
6. A. P. Potylitsyn, Phys. Lett. A **238**, 112 (1998).
7. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media*, 3rd ed. (Nauka, Moscow, 1992; Pergamon Press, Oxford, 1984).
8. V. E. Pafomov, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **10**, 240 (1967).
9. G. Kube, H. Backe, H. Euteneuer, *et al.*, Phys. Rev. E **65**, 056501 (2002).
10. V. P. Shestopalov, *Diffraction Electronics* (Khark. Gos. Univ, Kharkov, 1976) [in Russian].
11. J. H. Brownell, J. Walsh, H. G. Kirk, *et al.*, Nucl. Instrum. Methods Phys. Res. A **393**, 323 (1997).
12. Y. Shibata, S. Hasebe, K. Ishi, *et al.*, Phys. Rev. E **57**, 1061 (1998).

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 99, No. 2, 2004, pp. 320–327.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 2, 2004, pp. 359–368.

ATOMS, SPECTRA, RADIATION

Ionization and Stabilization of a Two-Electron Atom in a Strong Electromagnetic Field

E. A. Volkova, V. V. Gridchin, A. M. Popov, and O. V. Tikhonova

Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, 119992 Russia

e-mail: popov@mics.msu.su

Received March 3, 2004

Abstract—Direct numerical simulations are performed to analyze stabilization of a two-electron model atom in a strong electromagnetic field. The system is found to be stabilized with respect to both single and double ionization. By comparing the present results with those concerning stability of one-electron atoms, it is shown that stabilization is due to the formation of a Kramers–Henneberger two-electron atom. Ionization and stabilization characteristics of excited singlet and triplet states of an atomic system are examined. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Ionization of atoms in strong electromagnetic fields has been the subject of extensive theoretical and experimental research over the last 20 years (e.g., see [1–3]). One of the most interesting phenomena recently discovered is direct double (multielectron) photoionization, i.e., simultaneous release of two electrons from an atom in a strong electromagnetic field [4]. The resulting photoionization yield is substantially (sometimes, by orders of magnitude) higher than that predicted by a model that assumes that electrons are successively and independently removed from an atom in a laser field. A detailed survey of current studies of multielectron photoionization can be found in [5]. However, it is also known that an atomic system placed in a strong electromagnetic field is characterized by relatively higher stability. The basic mechanisms of this effect were originally considered in [6, 7] and discussed in detail in [8, 9]. Basically, it is attributed to a substantial change in the energy spectrum of an atomic system placed in a strong electromagnetic field and formation of a "dressed" atom. In a high-frequency laser field, a dressed atom behaves like a Kramers–Henneberger (KH) atom (see [9], where both energy spectrum and steady-state wavefunctions of the KH atom are discussed in detail).

It should be noted here that the stabilization effect in question has so far been investigated, both theoretically and experimentally, in the one-electron approximation. The first attempts to analyze the stabilization of a multielectron atom were made in [10, 11]. Those studies were focused on the KH potential of the hydrogen ion H^- placed in a strong high-frequency field. It was shown that the binding energy for this system increases with field intensity, and additional two-electron bound states arise. Moreover, the doubly charged negative hydrogen ion H^{2-} can exist. A stabilization regime analogous to stabilization of an unexcited one-electron atom was also predicted for extremely strong fields that give rise to a pronounced dichotomous structure of the KH potential. Similar results were obtained in [12] for the helium atom. In [13], stabilization was revealed by analyzing ionization dynamics for a model two-electron negative hydrogen ion in a high-frequency field (with photon energy $\hbar\omega$ exceeding the energy required to release both electrons), but the underlying physical mechanism was not investigated.

Thus, both conditions leading to stabilization of a multielectron atomic system and physical mechanisms responsible for it remain unclear to this day. In particular, the following problems should be addressed:

I. Determine the frequency ranges and intensities of stabilizing fields.

II. Explore the possibility of simultaneous decrease in single- and double-ionization probabilities with increasing field intensity.

III. Find and specify the difference (if any) in double-ionization and stabilization characteristics between singlet and triplet states.

These problems are analyzed in the present study.

2. MODEL

As a model atom, we consider a two-electron one-dimensional quantum-mechanical system described by the Hamiltonian (in atomic units)

$$H_0 = \sum_{i=1}^{2}(T_i + V(x_i)) + V_{12}(x_1, x_2), \qquad (1)$$

where T_i is the kinetic energy of the ith electron,

$$V(x_i) = -\frac{Z}{\sqrt{\alpha^2 + x_i^2}}$$

is the energy of its interaction with the nucleus,

$$V_{12} = \frac{1}{\sqrt{\alpha^2 + (x_1 - x_2)^2}}$$

is the energy of interaction between the electrons, $Z = 2$ is the atomic number, and $\alpha = 0.92$ Å is a smoothing parameter. The energy of the ground state $|1, 1\rangle$ is $E_0 = -37.3$ eV, and the single-ionization potential is 12.95 eV. This system was considered as a two-electron model of the xenon atom in an analysis of double-photoionization mechanisms [14].

The table lists the energies of several low eigenstates of the system. Figures 1 and 2 illustrate the electron density distributions in the ground state $|1, 1\rangle$ and in the lowest singlet and triplet states, $|1, 2\rangle_s$ and $|1, 2\rangle_t$, respectively.

Interaction between the system and an electromagnetic wave is described in the dipole approximation:

$$W(x_1, x_2, t) = (x_1 + x_2)\varepsilon(t)\cos\omega t, \qquad (2)$$

where ω is the electric field frequency and $\varepsilon(t)$ is the pulse envelope.

The time-dependent Schrödinger equation with the Hamiltonian

$$H = H_0(x_1, x_2) + W(x_1, x_2, t) \qquad (3)$$

was computed by applying the technique developed in [15] on a spatial grid in the coordinates

$$\xi = \frac{x_1 + x_2}{\sqrt{2}}, \quad \eta = \frac{x_1 - x_2}{\sqrt{2}},$$

where

$$\xi \in (-100\ \text{Å}, 100\ \text{Å}), \quad \eta \in (-100\ \text{Å}, 100\ \text{Å}).$$

The coordinates ξ and η are equal (up to a numerical factor), respectively, to the center-of-mass and relative-position coordinates of the electrons, which are used below. To preclude reflection of probability flux from the domain boundaries, we added a complex-valued potential to (3).

Computations were performed for $\hbar\omega = 15.5$ and 46.5 eV. These frequencies are currently of interest, in particular, in view of rapid progress in free-electron laser technology. We used a smoothed trapezoidal pulse of electromagnetic field with rise time τ_f and "plateau" duration τ_{pl}. In the "high-frequency" case, we set

$$\tau_f = 60T, \quad \tau_{pl} = 150T;$$

the "low" frequency was combined with

$$\tau_f = 20T, \quad \tau_{pl} = 50T.$$

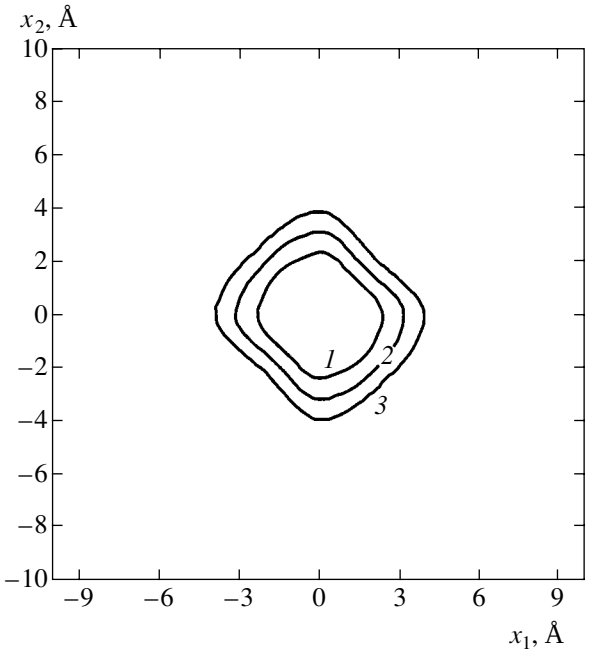


Fig. 1. Contour plot of electron density in the ground state $|1, 1\rangle$: 0.1 (*1*); 0,01 (*2*); 0.001 (*3*).

Here, $T = 2\pi/\omega$ is the wave period. These parameter values correspond to pulses of equal duration. Note that the energy of a high-frequency photon is sufficient to remove two atomic electrons simultaneously, whereas a low-frequency photon can remove only one electron and two photons are required to remove both electrons.

3. RESULTS

First, we discuss the results of computations of ionization dynamics for a ground-state two-electron atom interacting with the high-frequency field ($\hbar\omega = 46.5$ eV). Figure 3 shows the single- and double-ionization probabilities per pulse as functions of field intensity. These curves demonstrate that stabilization takes place when the field intensity P exceeds $P^* = 5 \times 10^{17}$ W/cm^2. Since the field frequency is relatively high, electron motion can be computed in nonrelativistic approximation in this intensity range.[1] Note that the

Energies of low singlet and triplet states

State	Energy, eV	
	singlet	triplet
$\|1, 1\rangle$	–37.3	–
$\|1, 2\rangle$	–29.5	–31.5
$\|1, 3\rangle$	–28.0	–28.3
$\|1, 4\rangle$	–26.7	–27.0
$\|1, 5\rangle$	–26.2	–26.3

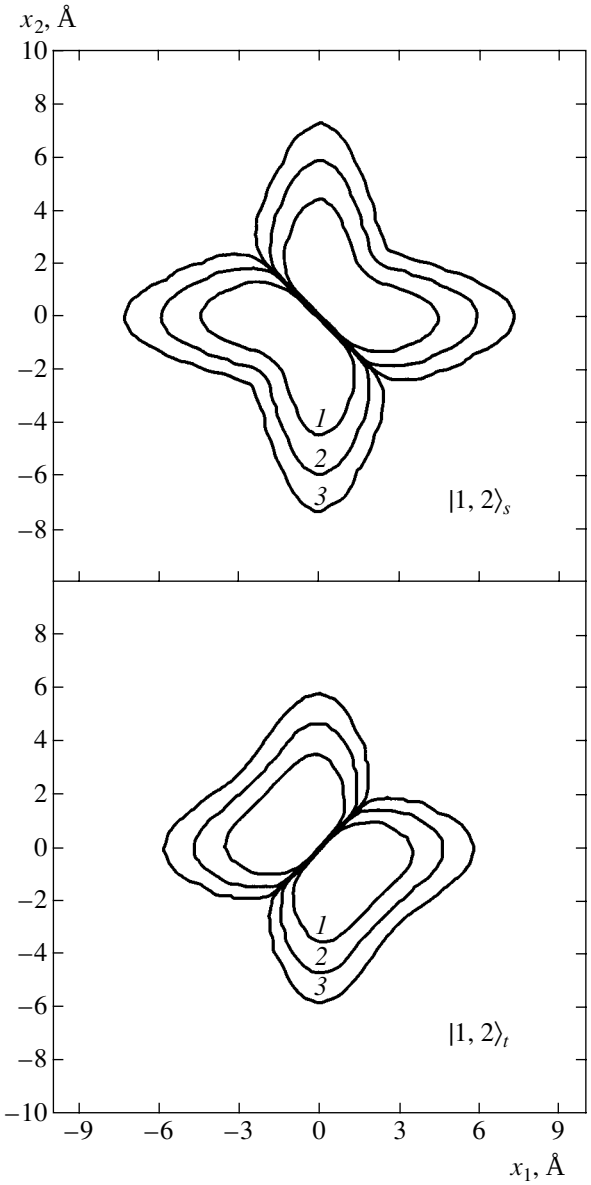


Fig. 2. Contour plots of electron density in the lowest excited states: singlet $|1, 2\rangle_s$ and triplet $|1, 2\rangle_t$.

atomic system is stabilized with respect to both single and double ionization (i.e., both probabilities reach maximum values) at the same intensity. Additional evidence of the system's stability with respect to ionization at high field intensities is provided by Fig. 4, which shows the time dependence of the probability of finding both electrons in the computational domain $V = \{\xi, \eta \in (-100\ \text{Å}, 100\ \text{Å})\}$,

$$W(t) = \iint_V |\psi(\xi, \eta, t)|^2 d\xi d\eta.$$

[1] For the frequency used in this example, relativistic effects are important when the field intensity is at least 10^{20} W/cm^2.

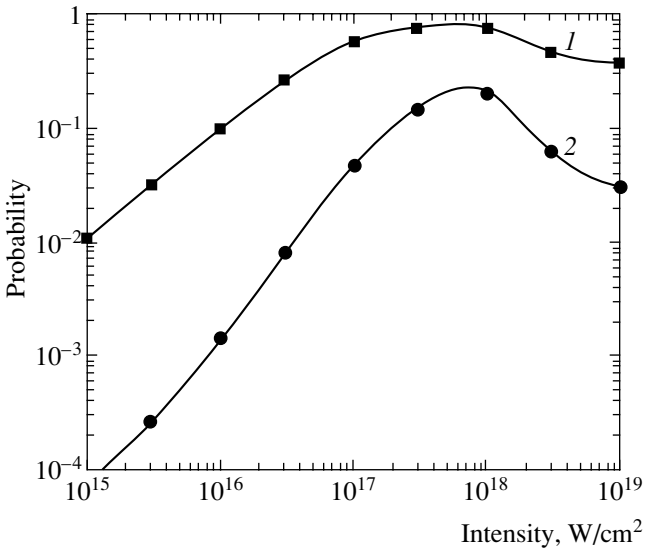


Fig. 3. Probabilities of single (*1*) and double (*2*) ionization per pulse versus field intensity for $\hbar\omega = 46.5$ eV.

When the field intensity is lower than the stabilization threshold, the steepest decrease in W with time, i.e., the highest ionization rate, corresponds to the "plateau" portion of the pulse (curve *1*). When the intensity exceeds the stabilization threshold, the steepest decrease in W corresponds to the rise and fall portions of the pulse, whereas the atom is relatively stable with respect to ionization within the plateau interval (curve *2*).

The stabilization demonstrated above can be explained by using the Kramers–Henneberger method. However, it should be kept in mind that the amplitude of free-electron oscillation in the field of an electromagnetic wave, $a_e = \varepsilon/\omega^2$, is comparable to the size of the ground-state wavefunction localization domain, $a \approx 2$ Å, under threshold conditions. Therefore, the dichotomous structure of the KH potential has no effect on stabilization (the KH potential is similar to the unperturbed atomic potential). Note that a detailed analysis of an analogous stabilization regime was performed in [16] for a one-electron system in a high-frequency field (with $\hbar\omega > I_i$, where I_i is ionization potential). In particular, it was shown that the stabilization was due to the nonlinearity of the matrix element corresponding to the KH harmonics responsible for bound–free transitions. Simple analytical estimates demonstrate a similarity to the example considered here. Indeed, when the field is relatively weak ($a_e \le a$), the bound-state and continuum wavefunctions for the KH potential are similar to the corresponding wavefunctions of a free atom. As an example, Fig. 5 shows the overlap integral $\langle\psi_{KH}|\psi_{at}\rangle$ for the ground-state wavefunctions corresponding to the atomic and KH potentials as a function of field intensity. When $P < 10^{18}$ W/cm^2 (which corresponds to $a_e \le a$), the value of $\langle\psi_{KH}|\psi_{at}\rangle$ is at least 0.9. In this case, a qualitative estimate for ionization proba-

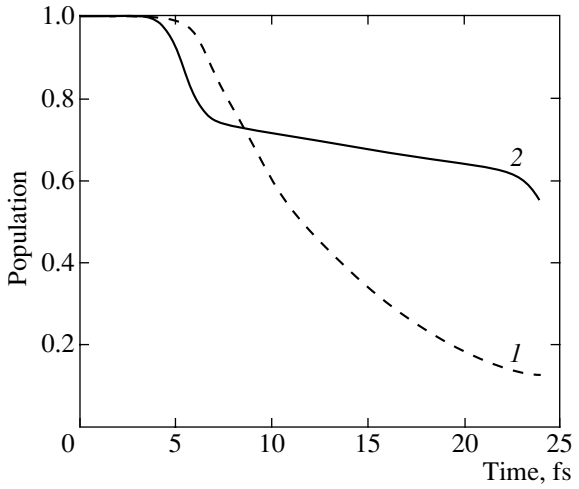


Fig. 4. Probability of finding both electrons in the computational domain for fields with $\hbar\omega = 46.5$ eV and intensities 3×10^{17} W/cm^2 (*1*) and 3×10^{18} W/cm^2 (*2*).

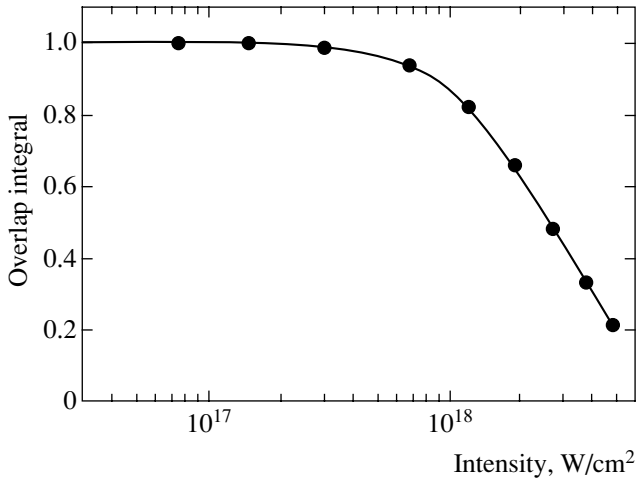


Fig. 5. Overlap integral versus intensity for ground states in atomic potential and corresponding KH potential.

bility can be obtained by evaluating the matrix element for the first KH harmonic,

$$V_1(x, X) = \frac{1}{2\pi}\int_0^{2\pi} V(x, X - a_e\cos\omega t)\exp(i\omega t)dt, \quad (4)$$

between unperturbed atomic wavefunctions:[2]

$$W_i \sim |\langle 1, 1|V_t|1, E\rangle|^2. \quad (5)$$

(Here, $X = (x_1 + x_2)/2$ and $x = x_1 - x_2$ are the center-of-mass and relative-position coordinates of the electrons, respectively, and V is the potential of interaction of the electrons with the nucleus and with one another.)

Curve *1* in Fig. 6 represents the squared absolute value of matrix element (5) calculated numerically as a function of field intensity. When the field is relatively weak, the ionization probability for a KH atom is a linear function of laser intensity corresponding to the first-order perturbation in the atom–field interaction. The absolute value of the matrix element reaches a maximum when the field intensity is such that $a_e \approx a$. With further increase in intensity, the value of $|\langle 1, 1|V_1|1, E\rangle|^2$ decreases, i.e., stabilization with respect to single ionization is observed. Curve *2* in Fig. 6 represents the squared absolute value of the matrix element evaluated by using the series expansion of $V_1(x, X)$ to the third order in a_e/x. A comparison of the curves demonstrates that the stabilization of the system is due to the nonlinear dependence of the bound–free transition matrix element on intensity. Moreover, a qualitative description of the phenomenon can be obtained by taking into account only one nonlinear term. It was pointed out in [17] that KH stabilization can be interpreted under these conditions in the basis of an unperturbed quantum-mechanical system as a result of interference between direct one-photon bound–free transitions and three–photon (multiphoton) transitions involving intermediate continuum states.

Next, we discuss the results concerning the dynamics of single and double ionization of the ground-state atom by the "low-frequency" field ($\hbar\omega = 15.5$ eV). In this case, single ionization is a one-photon process,

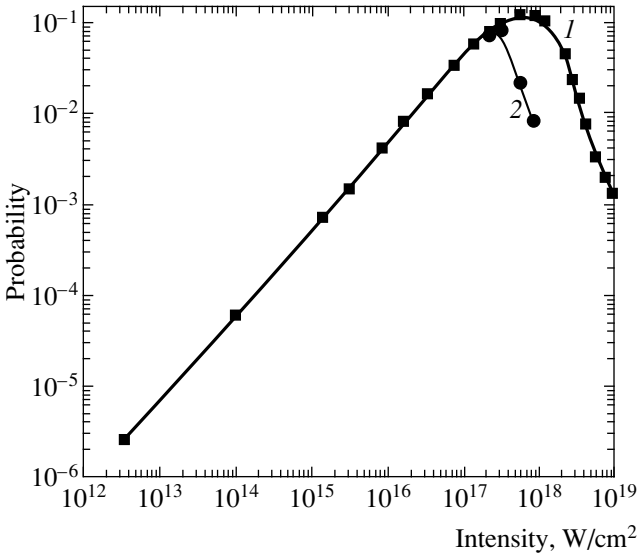


Fig. 6. Squared absolute value of matrix element for the first KH harmonic describing one-photon ionization (in arbitrary units): exact calculation (*1*) and expansion to the third order in a_e/x (*2*).

[2] The one-electron continuum wavefunction $|1, E\rangle$ was prescribed as the symmetrized product of the unperturbed one-electron bound-state function with the one-electron continuum wavefunction approximated by a plane wave with the wave vector $k = \sqrt{2mE/\hbar^2}$.

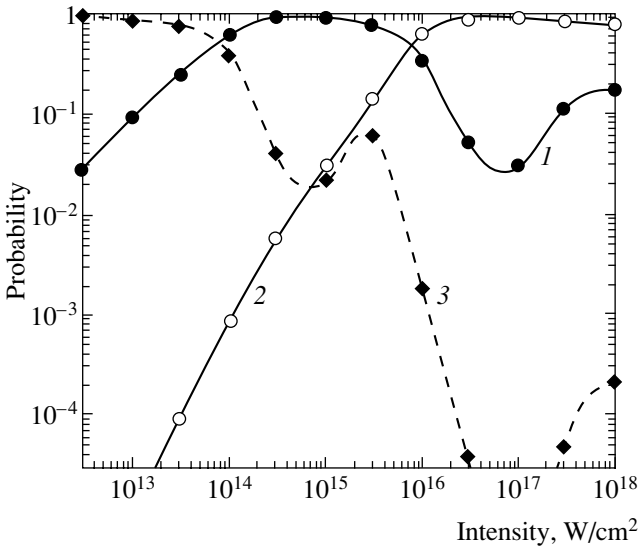


Fig. 7. Probabilities of single ionization (*1*), double ionization (*2*), and residual bound state (*3*) versus field intensity for $\hbar\omega = 15.5$ eV.

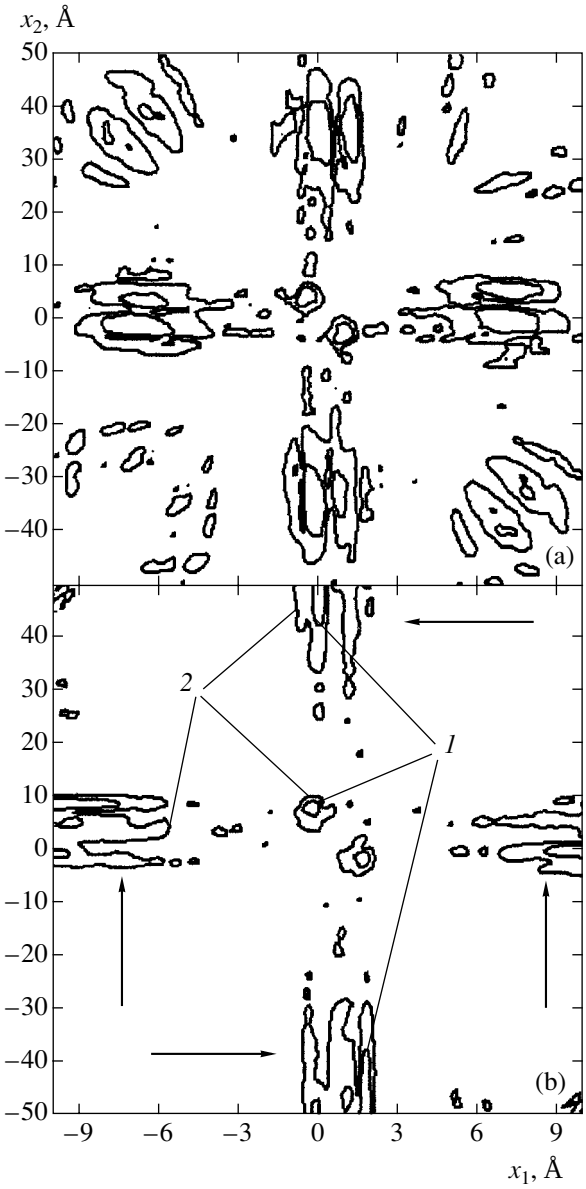


Fig. 8. Contour plots of electron density $|\psi(x_1, x_2)|^2$ obtained by numerical integration of the Schrödinger equation for $\hbar\omega = 15.5$ eV and a field intensity of 10^{18} W/cm^2 at the instants corresponding to (a) the end of the rise time and (b) the midpoint of the plateau portion of the pulse. The values of contour curves *1* and *2* differ by an order of magnitude. Arrows indicate the two-electron density corresponding to a singly charged KH ion.

whereas two electrons can be removed only by absorption of three photons. Accordingly, the system may approximately be treated as a one-electron atom when the field is relatively weak. This is confirmed by the single- and double-ionization probabilities per pulse plotted against field intensity in Fig. 7: double ionization can be neglected for intensities below 10^{15} W/cm^2. When the field intensity is on the order of 10^{15} W/cm^2, the probability that the atom remains in a bound state, presented in Fig. 7, corresponds to a stabilization regime characteristic of one-electron systems [9]. When the intensity exceeds 3×10^{15} W/cm^2, the contribution due to double ionization is substantial, and the stabilization regime breaks down. Note that the possibility of stabilization breakdown due to double photoionization was considered in [18]. According to our computations, further increase in intensity (to $P \geq 10^{17}$ W/cm^2) enhances stability of the system with respect to ionization. A detailed analysis of ionization dynamics shows that stabilization with respect to double ionization by high-intensity field is due to drastic transformation of the potential of an atomic system placed in a strong electromagnetic field and formation of a Kramers–Henneberger atom. Indeed, the electron density distributions $|\psi(x_1, x_2)|^2$ computed at different instants demonstrate the formation of a stable two-electron atom of diameter $2a_e = 2\varepsilon_0/\omega^2 \approx 10$ Å with a dichotomous wavefunction structure. The dichotomous structure is illustrated by the central portions of the distributions shown in Fig. 8. Under these conditions, the singly charged ion is a Kramers–Henneberger ion described by a polychotomous wavefunction (indicated by arrows in Fig. 8b). The Kramers–Henneberger two-electron atom is created when the barrier-suppression double-ionization threshold, $P_{\mathrm{BSI}} \approx 10^{17}$ W/cm^2, is exceeded. As the field intensity in the pulse falls below P_{BSI}, the KH wavefunction structure breaks down, and the ionization rate increases. Thus, a KH atom can be created in an electromagnetic field with photon energy lower than that required to remove both electrons only when the field intensity exceeds the barrier-suppression ionization threshold. This regime is analogous to the stabilization conditions obtained in [19] for a one-electron system in a low-frequency field (with $\hbar\omega < I_i$,

where I_i is ionization potential). As in the case considered in [19], a low residual bound-state population is obtained because electron density is "shaken off" the atom as the KH potential rapidly transforms into the potential of a free atom during the interaction with the trailing front of the pulse. The most likely result is one-electron continuum, which explains the increase in the probability of single ionization for intensities above the stabilization threshold (see Fig. 7).

Now, let us discuss the ionization and stabilization characteristics of singlet and triplet states of a two-electron system. To be specific, we consider photoionization of the $|1, 2\rangle_s$ and $|1, 2\rangle_t$ states, which are characterized by similar energies (see table) and the wavefunctions depicted in Fig. 2. Figure 9 shows the probabilities of single and double ionization by the high-frequency field computed as functions of field intensity for the singlet and triplet states. It is clear that both states of the system are stabilized with respect to single and double ionization. Moreover, the respective stabilization thresholds are equal to that for the ground state $|1, 1\rangle$. Stabilization is also predicted for field intensities that are not sufficiently high for the dichotomous structure of KH two-electron atomic states to develop.

It should be noted that the single-ionization probabilities are nearly equal for both states in the entire intensity range, whereas the triplet state is characterized by a higher double-ionization probability as compared to the singlet state, particularly at low intensities. Since both single and double ionization are one-photon processes at intensities up to 3×10^{17} W/cm^2 (the corresponding probabilities are linear functions of intensity), double ionization can occur only as a result of redistribution of the photon energy between the electrons. It is well known that the distance between electrons is greater in the triplet state as compared to the singlet state because of the difference in symmetry with respect to interchange of electrons between the singlet- and triplet-state wavefunctions. This leads to a difference in the rate of energy exchange between the electrons and, therefore, in the double-ionization probability.

To support this explanation, we estimate the ionization rates by applying perturbation theory and representing wavefunctions as combinations of one-electron orbitals:

$$|1, 2\rangle_{s(t)} \equiv \psi^{s(t)}_{1,2}(x_1, x_2) = \frac{1}{\sqrt{2}}(\psi_1(x_1)\psi_2(x_2) \pm \psi_2(x_1)\psi_1(x_2)) \quad (6)$$

for bound states,

$$|1, E\rangle_{s(t)} \equiv \psi^{s(t)}_{1,E}(x_1, x_2) = \frac{1}{\sqrt{2}}(\tilde{\psi}_1(x_1)\tilde{\psi}_E(x_2) \pm \tilde{\psi}_E(x_1)\tilde{\psi}_1(x_2)) \quad (7)$$

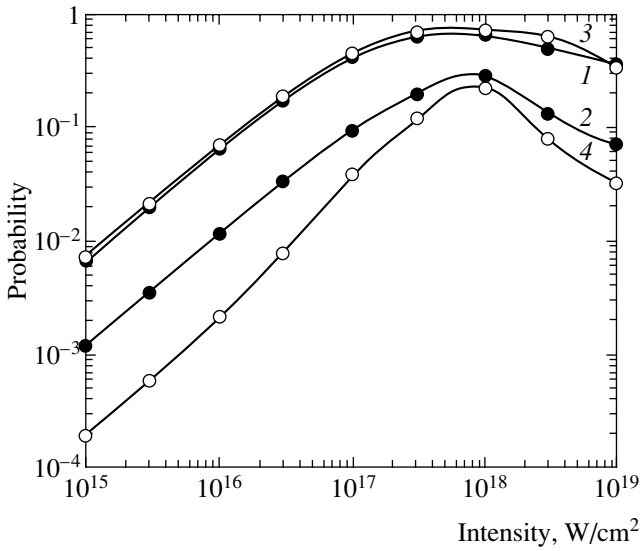


Fig. 9. Probabilities of single (*1*, *3*) and double (*2*, *4*) ionization of the $|1, 2\rangle_s$ (*1*, *2*) and $|1, 2\rangle_t$ (*3*, *4*) states versus field intensity for $\hbar\omega = 46.5$ eV.

for one-electron continuum, and

$$|E_1, E_2\rangle_{s(t)} \equiv \psi^{s(t)}_{E_1,E_2}(x_1, x_2) = \frac{1}{\sqrt{2}}(\tilde{\psi}_{E_1}(x_1)\tilde{\psi}_{E_2}(x_2) \pm \tilde{\psi}_{E_2}(x_1)\tilde{\psi}_{E_1}(x_2)) \quad (8)$$

for the two-electron continuum.

The one-electron wavefunctions used in (6)–(8) can be calculated in a self-consistent field approximation. Note that $\psi_1(x)$ and $\tilde{\psi}_1(x)$ are distinct, because the electrostatic potentials generated by the other electron in the states described by $\psi_2(x)$ and $\tilde{\psi}_E(x)$ are different. A similar distinction should be made between the one-electron continuum wavefunctions contained in (7) and (8).

The single-ionization probability amplitude is determined by the dipole matrix element

$$d^{1E}_{12} = \iint \psi^{s(t)}_{1,2}(x_1, x_2)(x_1 + x_2)\psi_{1,E}(x_1, x_2)dx_1dx_2. \quad (9)$$

Substituting (9) into (6) and (7), we obtain

$$(d^{1E}_{12})_{s,t} \sim d_{E,2}\langle\tilde{\psi}_1|\psi_1\rangle \pm d_{21}\langle\tilde{\psi}_E|\psi_1\rangle,$$

where

$$d_{E2} = \int \tilde{\psi}^*_E(x)x\psi_1(x)dx,$$

$$d_{21} = \int \tilde{\psi}^*_2(x)x\psi_1(x)dx$$

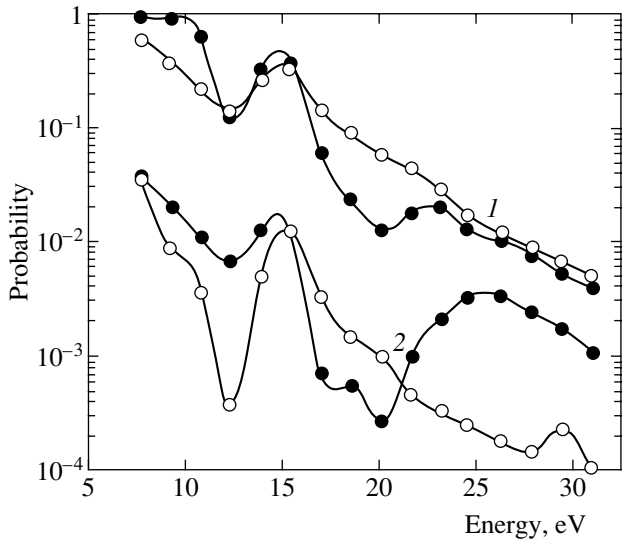


Fig. 10. Probabilities of single (*1*) and double (*2*) ionization of the $|1, 2\rangle_s$ (closed circles) and $|1, 2\rangle_t$ (open circles) states versus photon energy $\hbar\omega$ for a field intensity of 5×10^{13} W/cm^2. All pulse durations are equal, and their rise-time and plateau durations are multiples of the optical cycles.

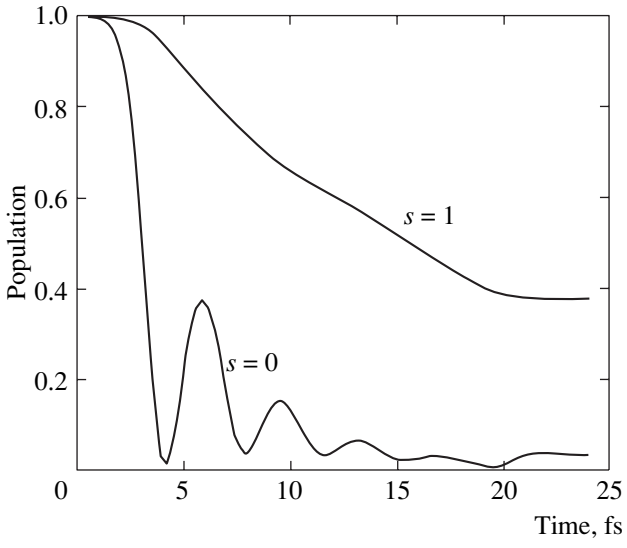


Fig. 11. Dynamics of decay of the $|1, 2\rangle_s$ (*1*, *2*) and $|1, 2\rangle_t$ states in the field with $\hbar\omega = 7.75$ eV and a field intensity of 5×10^{13} W/cm^2.

are one-electron dipole matrix elements and

$$\langle \tilde{\psi}_1 | \psi \rangle = \int \tilde{\psi}_1^*(x)\psi_1(x)dx,$$

$$\langle \tilde{\psi}_E | \psi \rangle = \int \tilde{\psi}_E^*(x)\psi_1(x)dx$$

are the overlap integrals for the ground-state and continuum wavefunctions contained in (6) and (7).

For a weakly correlated system,

$$\tilde{\psi}_1(x) \approx \psi_1(x).$$

In this case, we have

$$\langle \tilde{\psi}_1 | \psi \rangle \approx 1, \quad \langle \tilde{\psi}_E | \psi \rangle \ll 1$$

and the single- ionization probabilities for the singlet and triplet states are nearly equal, being determined by the matrix element d_{E2}.

To calculate the matrix element that determines the double-ionization probability amplitude, we write

$$d_{12}^{E_1E_2} = \iint (\psi_{E_1, E_2}^{s(t)}(x_1, x_2))^* \times (x_1 + x_2)\psi_{1,2}(x_1, x_2)dx_1dx_2. \quad (10)$$

Assuming that $E_1 \approx E_2 = E$ for symmetrized functions, we obtain

$$(d_{12}^{EE})_s \sim d_{E,1}\langle \tilde{\tilde{\psi}}_E | \psi_2 \rangle + \tilde{\tilde{\psi}}_{E,2} | \psi_1 \rangle;$$

in the case of antisymmetrized functions,

$$(d_{12}^{EE})_t \approx 0.$$

Thus, the one-photon double ionization, as treated in this approximation, is due to the fact that the ground-state basis functions are not orthogonal to continuum wavefunctions. This explains why the double-ionization probability is lower for triplet states as compared to singlet states.

Finally, we discuss ionization dynamics for the $|1, 2\rangle_s$ and $|1, 2\rangle_t$ states interacting with "low-frequency" fields ($\hbar\omega = 7.75$–30 eV). In this frequency range, single ionization is a one-photon process for both singlet and triplet states. However, a substantial difference in both single- and double-ionization probabilities between these states is predicted for $\hbar\omega \leq 25$ eV. Note also that a higher single-ionization probability is obtained either for the singlet state or for the triplet one, depending on field intensity (see Fig. 10). This is explained by the influence of autoionizing states in the energy range in question. Resonance between the initial bound state and an autoionizing state manifests itself by a nonmonotonic time dependence of the initial-state population during the interaction with a laser pulse (Rabi oscillation) and increases the rate of photoionization. The effect of resonance on the singlet state is demonstrated by comparing the dynamics of decay of the $|1, 2\rangle_s$ and $|1, 2\rangle_t$ states in the field with $\hbar\omega = 7.75$ eV illustrated by Fig. 11. It is clear that resonance leads to a higher decay rate for the singlet state as compared to the triplet one.[3] Depending on the field frequency, both

[3] The role played by autoionizing states in the double ionization of a ground-state two-electron atom was emphasized in [20].

singlet and triplet states may be in resonance with autoionizing states. This explains the nonmonotonic dependence of the ionization probabilities for the $|1, 2\rangle_s$ and $|1, 2\rangle_t$ states shown in Fig. 10 on photon energy. When $\hbar\omega \geq 25$ eV, no resonance with autoionizing states is observed and ionization characteristics are analogous to those predicted in the "high-frequency" case.

4. CONCLUSIONS

The present analysis of stabilization of a two-electron atom in a strong electromagnetic field shows that stability with respect to single and double ionization is enhanced as a certain threshold value of field intensity is exceeded. It is shown that the stabilization can be explained by using the Kramers–Henneberger method. When the field frequency exceeds the double-ionization threshold, the stabilization thresholds predicted for single and double ionization are equal. When the photon energy is not sufficiently high to ensure one-photon coupling between the ground state and two-electron continuum, but is higher than the single-ionization potential, different stabilization conditions are predicted with regard to single and double ionization. Stabilization with respect to double ionization corresponds to extremely high field intensities exceeding the barrier-suppression double-ionization threshold. Formation of the dichotomous structure of the two-electron wavefunction characteristic of a Kramers–Henneberger atom is graphically demonstrated. The considerable difference in double-ionization probability between singlet and triplet excited states of a two-electron atom is analyzed and explained. The important role played by autoionizing states in photoionization of a two-electron atom is demonstrated.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 03-02-16234, and by grant NSh-1713.200.2 under the Presidential Program of State Support of Leading Scientific Schools.

REFERENCES

1. M. V. Fedorov, *Atomic and Free Electrons in a Strong Light Field* (World Sci., Singapore, 1997).
2. N. B. Delone and V. P. Krainov, *Multiphoton Processes in Atoms*, 2nd ed. (Springer, Berlin, 2000).
3. N. B. Delone and V. P. Kraĭnov, *Nonlinear Ionization of Atoms by Laser Radiation* (Fizmatlit, Moscow, 2001) [in Russian].
4. A. L'Huiller, L. A. Lompre, G. Mainfray, and C. Manus, Phys. Rev. A **27**, 2503 (1983).
5. Opt. Express **8** (2001).
6. M. V. Fedorov and A. M. Movsesian, J. Phys. B **21**, L155 (1988).
7. M. Pont and M. Gavrila, Phys. Rev. Lett. **65**, 2362 (1990).
8. M. Gavrila, J. Phys. B **35**, R147 (2002).
9. A. M. Popov, O. V. Tikhonova, and E. A. Volkova, J. Phys. B **36**, R125 (2003).
10. H. G. Muller and M. Gavrila, Phys. Rev. Lett. **71**, 1693 (1993).
11. E. van Duijn, M. Gavrila, and H. G. Muller, Phys. Rev. Lett. **77**, 3759 (1996).
12. M. Gavrila and J. Shertzer, Phys. Rev. **53**, 3431 (1996).
13. E. A. Volkova, A. M. Popov, and O. V. Tikhonova, Zh. Éksp. Teor. Fiz. **122**, 978 (2002) [JETP **95**, 844 (2002)].
14. A. M. Popov, O. V. Tikhonova, and E. A. Volkova, Laser Phys. **11**, 236 (2001).
15. E. A. Volkova, A. M. Popov, and O. V. Tikhonova, Zh. Éksp. Teor. Fiz. **118**, 816 (2000) [JETP **91**, 706 (2000)].
16. A. M. Popov, O. V. Tikhonova, and E. A. Volkova, Laser Phys. **10**, 188 (2000).
17. O. V. Tikhonova, A. M. Popov, and M. V. Fedorov, Phys. Rev. A **65**, 053404 (2002).
18. E. A. Volkova, A. M. Popov, and O. V. Tikhonova, Zh. Éksp. Teor. Fiz. **114**, 1618 (1998) [JETP **87**, 875 (1998)].
19. A. M. Popov, O. V. Tikhonova, and E. A. Volkova, J. Phys. B **32**, 3331 (1999).
20. M. Sukharev, E. Charron, and A. Suzor-Weiner, Phys. Rev. A **66**, 053407 (2002).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 2, 2004, pp. 328–339.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 2, 2004, pp. 369–380.

PLASMA, GASES

On the Conductivity of a Magnetoactive Turbulent Plasma

O. G. Chkhetiani*

Space Research Institute, Russian Academy of Sciences, Profsoyuznaya ul. 84/32, Moscow, 117997 Russia
**e-mail: ochkheti@mx.iki.rssi.ru*

Received January 22, 2004

Abstract—The problem of determining the effective conductivity tensor of a magnetoactive turbulent plasma is considered in the approximation of isolated particles. Additional gyrotropic terms are shown to appear in the conductivity tensor in the presence of mean, nonzero magnetic helicity. The dispersion of propagating electromagnetic waves changes, additional modes and additional rotation of the polarization plane appear, and the waves can be amplified. The properties acquired by a plasma with helicity are similar to those observed in chiral and bianisotropic electrodynamic media. © *2004 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

An effective description of the propagation of waves and particles in fluctuational magnetic fields in a turbulent conductive medium is of great importance in solving various problems in plasma physics and astrophysics. The phenomena associated with the presence of small-scale magnetic helicity $\langle \mathbf{A} \cdot \mathbf{B} \rangle$ ($\mathbf{B} = \mathrm{curl}\,\mathbf{A}$), which manifest themselves virtually on all scales of plasma systems, play a special role here. Whereas large-scale helicity contributes to the stability of electromagnetic structures [1], its presence at the level of fluctuations is a nonequilibrium phenomenon that is accompanied by various large-scale instabilities [2]. Other effects produced by small-scale helicity, such as an asymmetry in the particle distribution and acceleration, are well known in the diffusion theory of cosmic-ray propagation [3–6]. The gyrotropic acceleration effects are also known in a laboratory plasma as helicity input effects [7, 8]. The appearance of additional helicity-related transport was also shown to be possible in [9]. Changes in the transport properties are also directly reflected in the dielectric (conductive) properties of a plasma medium. Thus, for example, it was shown in [10] that in the presence of fluctuational magnetic helicity in the low-conductivity limit, the effective current in an isotropic plasma proves to be dependent on the curl of the electric field ($\mathbf{j} = \sigma\mathbf{E} + \sigma_\kappa \mathrm{curl}\mathbf{E}$), which causes the mean magnetic field to grow under certain conditions. In [10], the external magnetic field was disregarded. In natural and laboratory conditions, the plasma is always under the influence of large-scale magnetic fields that significantly affect its properties [11]. Since magnetic helicity also emerges in plasma systems in the presence of a large-scale magnetic field, a study of its influence should take into account this factor. In [12], it was shown for an exactly solvable model of a nonlinear dynamo that the diffusion and generation rate are strongly suppressed even in a relatively weak magnetic field, and the regime of a fast dynamo transforms into the regime of a slow dynamo with a linear growth with time.

The goal of this work is to study the effective conductivity of a turbulent magnetoactive plasma with nonzero magnetic helicity. The kinetic approach is commonly used for a thorough theoretical description of plasma problems. However, allowance for the fluctuational effects of gyrotropy is rather difficult and is possible in finished form only with an appreciable number of assumptions and simplifications (see, e.g., [6, 9]). At the same time, many basic plasma properties can be determined in the approximation of isolated particles [11, 13], which will be used below.

The statistical parameters of the electromagnetic fluctuations are assumed to be stationary and uniform. In Section 2, we consider the equations of motion for particles and calculate the effective Lorentz force by the functional method with allowance for the nonuniformity of the electromagnetic perturbations to within the first order of perturbation theory. In Section 3, we determine the effective conductivity tensor. Fluctuational magnetic helicity gives rise to new gyrotropic terms. Our analysis of the dispersion relation both in the approximation of δ-correlated (in time) fluctuations (Section 4) and in the opposite case of long correlation times and high frequencies (Section 6) for electromagnetic waves and the evolution of the magnetic field (in the low-frequency limit) (Section 5) reveals changes in the dispersion of propagating waves and the presence of instabilities. The characteristic scales and growth rates of the instabilities are determined by the relationship between the fluctuational helicity and the energy and the external magnetic field. A magnetoactive turbulent plasma with helicity acquires properties similar to those of chiral and bianisotropic electrodynamic media, which have been extensively studied in recent years [14, 15]. In the Conclusions, we discuss our results and implications of the detected effect.

1063-7761/04/9902-0328$26.00 © 2004 MAIK "Nauka/Interperiodica"

2. BASIC EQUATIONS

Let us consider the motion of a one-component, singly charged plasma in a fluctuating electromagnetic field with given correlation properties. We will consider a cold plasma where the approximation of isolated particles [11, 13] can be used. A regular large-scale nonuniform perturbation of the electromagnetic field that is too weak to significantly change the correlation properties of the electromagnetic fluctuations, which are supposed to be given, stationary, and uniform, is assumed to arise in the system. The expression for the electron velocity $\mathbf{v}$ can be written as

$$\frac{d\mathbf{v}}{dt} = \frac{e}{m}\left(\mathbf{E} + \frac{1}{c}[\mathbf{v}\times\mathbf{B}]\right), \tag{1}$$

where e and m are the electron charge and mass, respectively. The electromagnetic field and the velocity can be represented as the sum of the large-scale slow component and the small-scale (with a zero mean) fast component:

$$\mathbf{E} = \langle\mathbf{E}\rangle + \tilde{\mathbf{E}}, \quad \mathbf{B} = \mathbf{B}_0 + \langle\mathbf{B}\rangle + \tilde{\mathbf{B}},$$
$$\mathbf{v} = \langle\mathbf{v}\rangle + \tilde{\mathbf{v}}.$$

As was said above, the mean electric and magnetic fields are assumed to be weak compared to the fluctuational fields, i.e., $\langle\mathbf{E}\rangle \ll \langle\tilde{\mathbf{E}}^2\rangle^{1/2}$ and $\langle\mathbf{B}\rangle \ll \langle\tilde{\mathbf{B}}^2\rangle^{1/2} < \mathbf{B}_0$. Passing to the Fourier representation, $\mathbf{F}(\mathbf{x}, t) = \int\hat{\mathbf{F}}(\mathbf{k}, w)\exp[i(\mathbf{k}\cdot\mathbf{x} - wt)]d\mathbf{k}dw$, we write

$$-iw\hat{\mathbf{v}}(\mathbf{k}, w) - \frac{e}{mc}[\hat{\mathbf{v}}(\mathbf{k}, w)\times\mathbf{B}_0] = \frac{e}{m}\hat{\mathbf{E}}(\mathbf{k}, w) + \frac{e}{mc}\int[\hat{\mathbf{v}}(\mathbf{q}, s)\times\hat{\mathbf{B}}(\mathbf{k}-\mathbf{q}, w-s)]d\mathbf{q}ds. \tag{2}$$

The equation of motion averaged over the electromagnetic fluctuations takes the form

$$-iw\langle\hat{\mathbf{v}}(\mathbf{k}, w)\rangle - \frac{e}{mc}[\langle\hat{\mathbf{v}}(\mathbf{k}, w)\rangle\times\mathbf{B}_0] = \frac{e}{m}\langle\hat{\mathbf{E}}(\mathbf{k}, w)\rangle + \frac{e}{mc}\int[\hat{\mathbf{v}}(\mathbf{q}, s) + \langle\hat{\mathbf{B}}(\mathbf{k}-\mathbf{q}, w-s)\rangle]d\mathbf{q}ds + \frac{e}{mc}\int\langle\hat{\tilde{\mathbf{v}}}(\mathbf{q}, s)\times\hat{\tilde{\mathbf{B}}}(\mathbf{k}-\mathbf{q}, w-s)\rangle d\mathbf{q}ds. \tag{3}$$

In view of the linear formulation of the problem, below we disregard the term

$$\frac{e}{mc}\int[\langle\hat{\mathbf{v}}(\mathbf{q}, s)\rangle\times\langle\hat{\mathbf{B}}(\mathbf{k}-\mathbf{q}, w-s)\rangle]d\mathbf{q}ds.$$

The correlation

$$\langle\hat{\tilde{\mathbf{v}}}(\mathbf{q}, s)\rangle\times\langle\hat{\tilde{\mathbf{B}}}(\mathbf{k}-\mathbf{q}, w-s)\rangle$$

can be expressed in terms of the cumulants of the fluctuational magnetic field by using the Furutsu–Novikov theorem [16]:

$$\langle\hat{\tilde{\mathbf{v}}}(\mathbf{q}, s)\rangle\times\langle\hat{\tilde{\mathbf{B}}}(\mathbf{k}-\mathbf{q}, w-s)\rangle_i = \varepsilon_{ijk}\int\left\langle\frac{\delta\hat{\tilde{V}}_j(\mathbf{q}, s)}{\delta\hat{\tilde{B}}_m(\mathbf{k}', w')}\right\rangle \times\langle\hat{\tilde{B}}_m(\mathbf{k}', w')\hat{\tilde{B}}_k(\mathbf{k}-\mathbf{q}, w-s)\rangle d\mathbf{k}'dw' + \varepsilon_{ijk}\int\left\langle\frac{\delta^2\hat{\tilde{V}}_j(\mathbf{q}, s)}{\delta\hat{\tilde{B}}_m(\mathbf{k}', w')\delta\hat{\tilde{B}}_n(\mathbf{k}'', w'')}\right\rangle\langle\hat{\tilde{B}}_m(\mathbf{k}', w')\hat{\tilde{B}}_n(\mathbf{k}'', w'') \times\hat{\tilde{B}}_k(\mathbf{k}-\mathbf{q}, w-s)\rangle d\mathbf{k}'d\mathbf{k}''dw'dw'' + \ldots, \tag{4}$$

where the variational derivative $\left\langle\frac{\delta\hat{\tilde{V}}_j(\mathbf{q}, s)}{\delta\hat{\tilde{B}}_m(\mathbf{k}', w')}\right\rangle$ satisfies the equation

$$\hat{L}_{js}(s)\left\langle\frac{\delta\hat{\tilde{V}}_s(\mathbf{q}, s)}{\delta\hat{\tilde{B}}_m(\mathbf{k}', w')}\right\rangle = -\frac{es}{mc}\varepsilon_{jlm}\frac{q_l}{q^2}\delta(s-w')\delta(\mathbf{q}-\mathbf{k}') + \frac{e}{mc}\varepsilon_{jlm}\langle\hat{V}_l(\mathbf{q}-\mathbf{k}', s-w')\rangle + \frac{e}{mc}\varepsilon_{jlr}\int\left\langle\frac{\delta\hat{\tilde{V}}_l(\mathbf{q}', s')}{\delta\hat{\tilde{B}}_m(\mathbf{k}', w')}\right\rangle\langle\hat{B}_r(\mathbf{q}-\mathbf{q}', s-s')\rangle d\mathbf{q}'ds' + \frac{e}{mc}\varepsilon_{jlr}\int\left\langle\frac{\delta^2\hat{\tilde{V}}_l(\mathbf{q}', s')}{\delta\hat{\tilde{B}}_m(\mathbf{k}', w')\delta\hat{\tilde{B}}_n(\mathbf{q}'-\mathbf{q}, s'-s)}\right\rangle \times\hat{Q}_{nr}(\mathbf{q}-\mathbf{q}', s-s')d\mathbf{q}'ds'. \tag{5}$$

Here,

$$\hat{L}_{js}(s) = -is\delta_{js} + \frac{e}{mc}\varepsilon_{jrs}B_{0r},$$
$$\langle\hat{\tilde{B}}_n(\mathbf{q}-\mathbf{q}', s-s')\hat{\tilde{B}}_r(\mathbf{k}, w)\rangle = \hat{Q}_{nr}(\mathbf{q}-\mathbf{q}', s-s')\delta(\mathbf{k}+\mathbf{q}-\mathbf{q}', w+s-s'). \tag{6}$$

The second variational derivative depends on the third derivative, etc. In general, the problem is not closed. In the case of δ-correlated (in time) fluctuations, the first term is retained in Eq. (4), which corresponds to the Gaussian approximation. This is also a good approximation for short correlation times. To take into account long correlation times, we can use, in particular, a consistent procedure of allowance for the memory effects

similar to that suggested in [17]. Having obtained the equation for the nth variational derivative, let us substitute the emerging term with the $(n+1)$th derivative with an effective relaxation term, which reflects the mixing role of the higher order moments. This, in turn, gives rise to an effective collision frequency determined by the pulsation amplitude of the magnetic field in the equation for the $(n-1)$th variational derivative, so the frequency s in an operator of the type $\hat{L}_{js}(s)$ changes to $s' \longrightarrow s + iw_*$. Here, we restrict our analysis to a simpler approach and set the last term in Eq. (5), as for a δ-correlated (in time) process, equal to zero. We can verify by direct analysis that this is possible when the characteristic frequencies of the electromagnetic fluctuations are much higher than the stochastic Larmor frequency determined from the mean amplitude of the magnetic fluctuations, $w_{\text{fluct}} \gg e\langle \tilde{\mathbf{B}}^2 \rangle^{1/2}/mc$. This approximation is similar to the "first post-Markovian" approximation used in the statistical theory of wave propagation in a turbulent medium [17]. Thus, for the first variational derivative, we write

$$\hat{L}_{js}(s)\left\langle \frac{\delta \hat{\hat{V}}_s(\mathbf{q}, s)}{\delta \hat{\hat{B}}_m(\mathbf{k}', w')} \right\rangle = -\frac{es}{mc}\varepsilon_{jlm}\frac{q_l}{q^2}\delta(s-w')\delta(\mathbf{q}-\mathbf{k}') + \frac{e}{mc}\varepsilon_{jlm}\langle \hat{V}_l(\mathbf{q}-\mathbf{k}, s-w')\rangle + \frac{e}{mc}\varepsilon_{jlr}\int\left\langle \frac{\delta \hat{\hat{V}}_l(\mathbf{q}', s')}{\delta \hat{\hat{B}}_m(\mathbf{k}', w')} \right\rangle \langle \hat{B}_r(\mathbf{q}-\mathbf{q}', s-s')\rangle d\mathbf{q}' ds'. \quad (7)$$

We take into account the nonuniformity of the mean field by successive approximations:

$$\hat{L}_{js}(s)\left\langle \frac{\delta \hat{\hat{V}}_s(\mathbf{q}, s)}{\delta \hat{\hat{B}}_m(\mathbf{k}', w')} \right\rangle^{(0)} = -\frac{es}{mc}\varepsilon_{jlm}\frac{q_l}{q^2}\delta(s-w')\delta(\mathbf{q}-\mathbf{k}') + \frac{e}{mc}\varepsilon_{jlm}\langle \hat{V}_l(\mathbf{q}-\mathbf{k}', s-w')\rangle, \quad (8)$$

$$\hat{L}_{js}(s')\left\langle \frac{\delta \hat{\hat{V}}_s(\mathbf{q}, s)}{\delta \hat{\hat{B}}_m(\mathbf{k}', w')} \right\rangle^{(1)} = \frac{e}{mc}\varepsilon_{jlr} \times \int\left\langle \frac{\delta \hat{\hat{V}}_l(\mathbf{q}', s')}{\delta \hat{\hat{B}}_m(\mathbf{k}', w')} \right\rangle^{(0)} \langle \hat{B}_r(\mathbf{q}-\mathbf{q}', s-s')\rangle d\mathbf{q}' ds'. \quad (9)$$

Retaining only the linear terms, we write

$$\left\langle \frac{\delta \hat{\hat{V}}_j(\mathbf{q}, s)}{\delta \hat{\hat{B}}_m(k', w')} \right\rangle = \hat{L}_{js}^{-1}(s) \times \left(-\frac{es}{mc}\varepsilon_{spm}\frac{q_p}{q^2}\delta(s-w')\delta(\mathbf{q}-\mathbf{k}') + \frac{e}{mc}\varepsilon_{spm}\langle \hat{V}_p(\mathbf{q}-\mathbf{k}', s-w')\rangle\right) - \left(\frac{e}{mc}\right)^2\varepsilon_{jlr}\varepsilon_{tpm} \times \hat{L}_{js}^{-1}(s)\hat{L}_{lt}^{-1}(w')w'\frac{k_p'}{k'^2}\langle \hat{B}_r(\mathbf{q}-\mathbf{k}', s-w')\rangle. \quad (10)$$

Here, $\hat{L}_{ij}^{-1}(s)$ is the operator that is the inverse of $\hat{L}_{ij}(s)$:

$$\hat{L}_{ij}^{-1}(s) = \frac{1}{is(s^2-\Omega_e^2)} \times (-\delta_{ij}s^2 + \Omega_{ei}\Omega_{ej} - is\varepsilon_{ijk}\Omega_{ek}), \quad \boldsymbol{\Omega}_e = \frac{e\mathbf{B}_0}{mc}. \quad (11)$$

In what follows, we use the relationship between the fields **B** and **E** via Maxwell's equation written in the Fourier representation as

$$\hat{\mathbf{B}} = \frac{c}{w}[\mathbf{k}\times\hat{\mathbf{E}}]. \quad (12)$$

For uniform gyrotropic fluctuations with the anisotropy introduced by a uniform magnetic field, the correlation tensor $\hat{Q}_{mk}(\mathbf{q}, s)$ is [18–20]

$$\hat{Q}_{mk}(\mathbf{q}, s) = \left(\delta_{mk} - \frac{q_m q_k}{q^2}\right)\frac{E_M(q, \mathbf{l}\cdot\mathbf{q}, s)}{4\pi q^2} + i\frac{H_M(q, \mathbf{l}\cdot\mathbf{q}, s)}{8\pi q^4}\varepsilon_{mkt}q_t + \left[(l_m q_k + l_k q_m)(\mathbf{l}\cdot\mathbf{q}) - l_m l_k q^2 - \frac{q_m q_k}{q^2}(\mathbf{l}\cdot\mathbf{q})^2\right] \times \frac{F(q, \mathbf{l}\cdot\mathbf{q}, s)}{4\pi q^4} - i(\delta_{ml}\varepsilon_{kij} + \delta_{kl}\varepsilon_{mij}) \times l_i l_j (l_l q^2 - q_l(\mathbf{l}\cdot\mathbf{q}))\frac{C(q, \mathbf{l}\cdot\mathbf{q}, s)}{4\pi q^4}. \quad (13)$$

Here, **l** is a unit vector parallel to the uniform magnetic field, $\mathbf{l} \parallel \mathbf{B}_0$. All of the correlation functions, except for $C(q, \mathbf{l}\cdot\mathbf{q}, s)$, are even in $\mathbf{l}\cdot\mathbf{q}$. The symmetry properties also admit combinations linear in components of the vector **l** considered in [6, 9]. However, it was shown in [18, 20] that when the anisotropy is attributable to a

magnetic field, the only possible combinations are quadratic ones.[1] This is also confirmed by direct calculations of the magnetic field effect on the correlation properties of turbulence [21]. For a weak anisotropy (and for obtaining analytical results), we can use the representation

$$\begin{aligned} E_M(q, \mathbf{l}\cdot\mathbf{q}, s) &= E(q, s) - \frac{(\mathbf{l}\cdot\mathbf{q})^2}{q^2}E_1(q, s), \\ H_M(q, \mathbf{l}\cdot\mathbf{q}, s) &= H(q, s) - \frac{(\mathbf{l}\cdot\mathbf{q})^2}{q^2}H_1(q, s), \\ F(q, \mathbf{l}\cdot\mathbf{q}, s) &= F(q, s), \\ C(q, \mathbf{l}\cdot\mathbf{q}, s) &= C_1(q, s)(\mathbf{l}\cdot\mathbf{q}). \end{aligned} \tag{14}$$

Assuming the decay of the correlations with time to be exponential, $\sim(\tau_*/\tau)\exp(-|t - t'|/\tau)$, we write for the Fourier transform

$$f(q, s) = f(q)\frac{\tau_*}{\pi(1 + s^2\tau^2)}. \tag{15}$$

Here, τ_* is the time constant determined by the characteristic frequencies and scales. Thus, for example, for interplanetary plasma turbulence [22], τ_* is assumed to be

$$\tau_* \sim \frac{\lambda}{v_A} = \frac{\lambda\omega_i}{c\Omega_i},$$

where λ is the characteristic fluctuational scale of the magnetic nonuniformities. Clearly, this estimate may also be valid for ionospheric plasma.

Let us expand the tensor $Q_{mk}(\mathbf{k} - \mathbf{q}, w - s) = \hat{Q}_{km}(\mathbf{q} - \mathbf{k}, s - w)$ as a series in $k \ll q$,

$$\begin{aligned} \hat{Q}_{km}(\mathbf{q} - \mathbf{k}, s - w) &= \hat{Q}_{km}(\mathbf{q}, s) \\ - k_r\frac{\partial\hat{Q}_{km}(\mathbf{q}, s)}{\partial q_r} &+ \frac{k_r k_t}{2}\frac{\partial\hat{Q}_{km}(\mathbf{q}, s)}{\partial q_r\partial q_t} + \ldots, \end{aligned} \tag{16}$$

and substitute this representation in (10), performing the integration over the solid angles, the frequencies s. We then find that, to within the first degree of the expansion in terms of the correlation time τ and neglecting the effects quadratic in wave vector ($\sim k^2$), the Lorentz force averaged over the uniform electromagnetic background fluctuations is

$$\begin{aligned} &\frac{e}{mc}\int\langle\hat{\tilde{\mathbf{v}}}(\mathbf{q}, s)\times\hat{\tilde{\mathbf{B}}}(\mathbf{k} - \mathbf{q}, w - s)\rangle d\mathbf{q}ds \\ &= -\left(\frac{e}{mc}\right)^2\hat{\mathscr{E}}\tau_*\langle\hat{\mathbf{v}}(\mathbf{k}, w)\rangle \\ &+ \frac{e}{m}\left(\frac{e}{mc}\right)^2\hat{\mathscr{H}}\tau_* i[\mathbf{k}\times\langle\hat{\mathbf{E}}(\mathbf{k}, w)\rangle] + \frac{2}{3}\left(\frac{e}{mc}\right)^2 \\ &\times E_0\tau_*\tau\left(1 - \frac{1}{5}t_1 + \frac{4}{5}t_2\right)[\boldsymbol{\Omega}_e\times\langle\hat{\mathbf{v}}(\mathbf{k}, w)\rangle] \\ &- \frac{2}{3}\frac{e}{m}\left(\frac{e}{mc}\right)^2\frac{H_0\tau\tau_*}{w}\left(1 - \frac{3}{10}g_1\right) \\ &\times[\boldsymbol{\Omega}_e\times[\mathbf{k}\times\langle\hat{\mathbf{E}}(\mathbf{k}, w)\rangle]] \\ &- \frac{2}{3}\left(\frac{e}{mc}\right)^2 H_0\tau_*(1 + i\tau w)\boldsymbol{\Omega}_e\delta(\mathbf{k})\delta(w), \end{aligned}$$

where

$$[\hat{\mathscr{H}}]_{ij} = \mathscr{H}_\perp\delta_{ij} + (\mathscr{H}_\parallel - \mathscr{H}_\perp)l_i l_j, \tag{17}$$

$$\begin{aligned} \mathscr{H}_\perp &= H_0\frac{2i}{3w}(1 + iw\tau)\left(1 - \frac{3}{10}g_1\right), \\ \mathscr{H}_\parallel &= H_0\frac{2i}{3w}(1 + iw\tau)\left(1 - \frac{2}{5}g_1\right), \end{aligned} \tag{18}$$

$$[\hat{\mathscr{E}}_{ij}] = \mathscr{E}_\perp\delta_{ij} + (\mathscr{E}_\parallel - \mathscr{E}_\perp)l_i l_j,$$

$$\mathscr{E}_\perp = \frac{4}{3}E_0(1 + i\tau w)\left(1 - \frac{3}{10}t_1 + \frac{9}{20}t_2\right), \tag{19}$$

$$\mathscr{E}_\parallel = \frac{4}{3}E_0(1 + i\tau w)\left(1 - \frac{2}{5}t_1 + \frac{1}{10}t_2\right). \tag{20}$$

Here,

$$\begin{aligned} H_0 &= \int\frac{H(q)}{q^2}dq = \langle\tilde{\mathbf{A}}\cdot\tilde{\mathbf{B}}\rangle_0, \\ H_1 &= \int\frac{H_1(q)}{q^2}dq, \end{aligned} \tag{21}$$

$$E_0 = \int E(q)dq = \langle\tilde{\mathbf{B}}^2\rangle_0, \quad E_1 = \int E_1(q)dq, \quad E_2 = \int F(q)dq, \tag{22}$$

$$g_1 = \frac{H_1}{H_0}, \quad t_1 = \frac{E_1}{E_0}, \quad t_2 = \frac{E_2}{E_0}, \quad q = |\mathbf{q}|. \tag{23}$$

The subscript 0 corresponds to the isotropic case. As we see, the effective transport coefficients are directly

[1] Indeed, an arbitrary vortex field can be represented as a sum of its toroidal and poloidal components with the basis defined for an arbitrary direction of $\mathbf{l}$:

$$h_i(\mathbf{x}) = l_k\frac{\partial^2 P}{\partial x_k\partial x_i} - l_i\Delta P + \varepsilon_{ikj}l_k\frac{\partial T}{\partial x_j}.$$

Choosing the direction of the external stationary uniform magnetic field as this direction, we find that the dependence on the components of this direction appears in the tensor of the pair correlations between the magnetic fluctuations only quadratically.

related to the mean energy and helicity of the fluctuational magnetic field.

For the time being, let us restrict our analysis to the approximation of a δ-correlated process, $\tau \longrightarrow 0$. The effects of finite correlation times will be considered below. For the average Lorentz force, we then obtain

$$\frac{e}{mc}\int \langle \hat{\hat{\mathbf{v}}}(\mathbf{q}, s) \times \hat{\hat{\mathbf{B}}}(\mathbf{k}-\mathbf{q}, w-s)\rangle d\mathbf{q}ds$$
$$= \frac{e}{m}\left(\frac{e}{mc}\right)^2 \hat{\mathcal{H}}_{\tau\to 0}\tau_1 i[\mathbf{k}\times\langle\hat{\mathbf{E}}(\mathbf{k}, w)\rangle] \qquad (24)$$
$$-\left(\frac{e}{mc}\right)^2 \hat{\mathcal{E}}_{\tau\to 0}\tau_1\langle\hat{\mathbf{v}}(\mathbf{k}, w)\rangle - \left(\frac{e}{mc}\right)^2\frac{2}{3}H_0\tau_1\boldsymbol{\Omega}_e.$$

The last term on the right-hand side of Eq. (24) has the meaning of constant acceleration along the external magnetic field. To all appearances, the possibility of such acceleration was first pointed out in [5] (see also [6]) and was also considered in detail in [7, 8] when the helicity input was discussed. It was suggested, as an explanation, that the acceleration is produced by the electric field generated by a fluctuational dynamo effect. Attention to the relationship between the acceleration effect and the transfer of electromagnetic field momentum to particles of the medium was drawn in [23]. Assuming that $\langle\mathbf{E}\rangle$ and $\langle\mathbf{B}\rangle$ are equal to zero, we find that in the nonrelativistic collisionless limit, a charged particle reaches a velocity

$$\mathbf{v}_{\max} \approx -\frac{1}{2}\frac{\langle\mathbf{A}\cdot\mathbf{B}\rangle}{\langle\mathbf{B}^2\rangle}\boldsymbol{\Omega}_e;$$

i.e., it does not depend on the correlation time and is determined by the Larmor frequency in external magnetic field and by the scale specified by the relationship between magnetic helicity and energy. In what follows, we disregard this effect. This is possible for $|\mathbf{k}\cdot\mathbf{v}_{\max}|/\omega_e \ll 1$, where $\omega_e^2 = 4\pi ne^2/m$, and n is the electron density.

3. THE CONDUCTIVITY TENSOR

Given the fluctuational friction specified by the term

$$-\left(\frac{e}{mc}\right)^2\hat{\mathcal{E}}_0\tau_*\langle\hat{\mathbf{v}}(\mathbf{k}, w)\rangle,$$

the inverse operator is

$$\hat{L}_{ij}^{-1}(w) = \frac{-\delta_{ij}(w+i\overline{\Omega_{\perp e}^2}\tau_*)^2 + \Omega_{ei}\Omega_{ej} - i(w+i\overline{\Omega_{\|e}^2}\tau_*)\varepsilon_{ijk}\Omega_{ek}}{i(w+i\overline{\Omega_{\|e}^2}\tau_*)((w+i\overline{\Omega_{\perp e}^2}\tau_*)^2-\Omega_e^2)}. \qquad (25)$$

Here,

$$\overline{\Omega_{\perp e}^2} = \left(\frac{e}{mc}\right)^2\mathcal{E}_{0\perp}, \quad \overline{\Omega_{\|_e}^2} = \left(\frac{e}{mc}\right)^2\mathcal{E}_{0\|}.$$

Taking into account the explicit form of the tensor $\hat{\mathcal{H}}_0$, let us write the electron velocity as

$$\langle\hat{\mathbf{v}}(\mathbf{k}, w)\rangle = -\frac{e}{m}\frac{\langle\hat{\mathbf{E}}(\mathbf{k}, w)\rangle}{iw'_{e\|}}$$
$$+\frac{e}{m}\frac{\Omega_e^2(\mathbf{l}(\mathbf{l}\cdot\langle\hat{\mathbf{E}}(\mathbf{k}, w)\rangle)-\langle\hat{\mathbf{E}}(\mathbf{k}, w)\rangle)}{iw'_{e\|}(w'^2_{e\perp}-\Omega_e^2)}$$
$$+\frac{e}{m}\frac{\Omega_e[\mathbf{l}\times\langle\hat{\mathbf{E}}(\mathbf{k}, w)\rangle]}{w'^2_{e\perp}-\Omega_e^2} - \frac{e}{m}\left(\frac{e}{mc}\right)^2\mathcal{H}_{0\perp}\tau_* \qquad (26)$$
$$\times\frac{i}{w}\frac{w'^2_{e\perp}(\mathbf{k}\times\mathbf{E}(\mathbf{k})-\mathbf{l}(\mathbf{l}\cdot[\mathbf{k}\times\langle\hat{\mathbf{E}}(\mathbf{k}, w)\rangle]))}{w'_{e\|}(w'^2_{e\perp}-\Omega_e^2)}$$
$$-\frac{e}{m}\left(\frac{e}{mc}\right)^2\mathcal{H}_{0\|}\tau_*\frac{i}{w}\frac{\mathbf{l}(\mathbf{l}\cdot[\mathbf{k}\times\langle\hat{\mathbf{E}}(\mathbf{k}, w)\rangle])}{w'_{e\|}}$$
$$-\frac{e}{m}\left(\frac{e}{mc}\right)^2\mathcal{H}_{0\perp}\tau_*\Omega_e\frac{\mathbf{l}\times[\mathbf{k}\times\langle\hat{\mathbf{E}}(\mathbf{k}, w)\rangle]}{w(w'^2_{e\perp}-\Omega_e^2)},$$

where

$$\Omega_e = \frac{e|\mathbf{B}_0|}{mc}, \quad w'_{e\perp(\|)} = w + i\overline{\Omega_{\perp(\|)}^2}\tau_*.$$

The calculations for ions are similar, and the ion velocity can be written as

$$\langle\hat{\mathbf{v}}(\mathbf{k}, w)\rangle_i = \frac{e}{M}\frac{\langle\hat{\mathbf{E}}(\mathbf{k}, w)\rangle}{iw'_{i\|}}$$
$$-\frac{e}{M}\frac{\Omega_i^2(\mathbf{l}(\mathbf{l}\cdot\langle\hat{\mathbf{E}}(\mathbf{k}, w)\rangle)-\langle\hat{\mathbf{E}}(\mathbf{k}, w)\rangle)}{iw'_{i\|}(w'^2_{i\perp}-\Omega_i^2)}$$

$$+\frac{e}{M}\frac{\Omega_i[\mathbf{l}\times\langle\hat{\mathbf{E}}(\mathbf{k},w)\rangle]}{(w'^2_{i\perp}-\Omega_i^2)}+\frac{e}{M}\left(\frac{e}{Mc}\right)^2\mathcal{H}_{0\perp}\tau_*$$
$$\times\frac{i}{w}\frac{w'^2_{i\perp}(\mathbf{k}\times\langle\hat{\mathbf{E}}(\mathbf{k},w)\rangle-\mathbf{l}(\mathbf{l}\cdot[\mathbf{k}\times\langle\hat{\mathbf{E}}(\mathbf{k},w)\rangle]))}{w'_{i\|}(w'^2_{i\perp}-\Omega_i^2)} \quad (27)$$
$$+\frac{e}{M}\left(\frac{e}{Mc}\right)^2\mathcal{H}_{0\|}\tau_*\frac{i}{w}\frac{\mathbf{l}(\mathbf{l}\cdot[\mathbf{k}\times\langle\hat{\mathbf{E}}(\mathbf{k},w)\rangle])}{w'_{i\|}}$$
$$-\frac{e}{M}\left(\frac{e}{Mc}\right)^2\mathcal{H}_{0\perp}\tau_*\Omega_i\frac{\mathbf{l}\times[\mathbf{k}\times\langle\hat{\mathbf{E}}(\mathbf{k},w)\rangle]}{w(w'^2_{i\perp}-\Omega_i^2)}.$$

The subscript i refers to the ion analogs of the parameters introduced for electrons:

$$\Omega_i=\frac{e|\mathbf{B}|_0}{Mc},\quad w'_{i\perp(\|)}=w+i\overline{\Omega^2_{\perp(\|)_i}}\tau_*.$$

As we see, averaging over the electromagnetic fluctuations is equivalent, in particular, to an effective collision with frequencies proportional to $\overline{\Omega^2_{\perp_{e(i)}}}\tau_*$ and $\overline{\Omega^2_{\|_{e(i)}}}\tau_*$.

For the conductivity tensor $j_k=\hat{\sigma}_{kl}(\mathbf{k},w)E_l(\mathbf{k},w)$ ($\mathbf{j}=ne(\langle\mathbf{v}\rangle-\langle\mathbf{v}\rangle_i)$), we obtain

$$4\pi\hat{\sigma}_{kl}(\mathbf{k},w)=-\left(\frac{\omega_e^2w'^2_{e\perp}}{iw'_{e\|}(w'^2_{e\perp}-\Omega_e^2)}+\frac{\omega_i^2w'^2_{i\perp}}{iw'_{i\|}(w'^2_{i\perp}-\Omega_i^2)}\right)\delta_{kl}$$
$$+\left(\frac{\omega_e^2\Omega_e^2}{iw'_{e\|}(w'^2_{e\perp}-\Omega_e^2)}+\frac{\omega_i^2\Omega_i^2}{iw'_{i\|}(w'^2_{i\perp}-\Omega_i^2)}\right)l_kl_l$$
$$+\left(\frac{\omega_e^2\Omega_e}{(w'^2_{e\perp}-\Omega_e^2)}-\frac{\omega_i^2\Omega_i}{(w'^2_{i\perp}-\Omega_i^2)}\right)\varepsilon_{kml}l_m$$
$$+\left(\frac{\omega_e^2h_{\perp e}\Omega_e}{w(w'^2_{e\perp}-\Omega_e^2)}+\frac{\omega_i^2h_{\perp i}\Omega_i}{w(w'^2_{i\perp}-\Omega_i^2)}\right)(l_mk_m\delta_{kl}-l_lk_k) \quad (28)$$
$$-i\left(\frac{\omega_e^2h_{\|e}}{ww'_{e\|}}+\frac{\omega_i^2h_{\|ei}}{ww'_{i\|}}\right)l_kl_m\varepsilon_{mnl}k_n$$
$$-i\left(\frac{\omega_e^2h_{\|e}w'_{e\perp}}{w(w'^2_{e\perp}-\Omega_e^2)}+\frac{\omega_i^2h_{\|ei}w'_{i\perp}}{w(w'^2_{i\perp}-\Omega_i^2)}\right)$$
$$\times(\varepsilon_{kml}k_m-l_kl_m\varepsilon_{lmn}k_n),$$

where $\omega^2_{e(i)}=4\pi ne^2/m(M)$.

The coefficients $h_{\perp e(i)}$ and $h_{\|e(i)}$ have the dimensions of velocity, and it is convenient to represent them as

$$h_{\perp e(i)}=\overline{\Omega^2_{\perp e(i)}}\tau_*\lambda_{k\perp}=\alpha_{\perp e(i)}\frac{\Omega_{e(i)}}{\Omega_{\kappa\perp}}c,$$
$$\alpha_{\perp e(i)}=\frac{\overline{\Omega^2_{\perp e(i)}}\tau_*}{\Omega_{e(i)}},\quad \Omega_{\kappa\perp}=\frac{c}{\lambda_{\kappa\perp}},$$
$$h_{\|e(i)}=\overline{\Omega^2_{\|e(i)}}\tau_*\lambda_{k\|}=\alpha_{\|e(i)}\frac{\Omega_{e(i)}}{\Omega_{\kappa\|}}c, \quad (29)$$
$$\alpha_{\|e(i)}=\frac{\overline{\Omega^2_{\perp e(i)}}\tau_*}{\Omega_{e(i)}},\quad \Omega_{\kappa\|}=\frac{c}{\lambda_{\kappa\|}},$$

where the scale $\lambda_{\kappa\perp(\|)}$ is defined by the ratio of the helicity and energy of the fluctuations:[2]

$$\lambda_{\kappa\perp(\|)}=\frac{\mathcal{H}_{0\perp(\|)}}{\mathcal{E}_{0\perp(\|)}}\approx\frac{1}{2}\frac{\langle\mathbf{A}\cdot\boldsymbol{B}\rangle}{\langle\mathbf{B}^2\rangle}. \quad (30)$$

Neglecting the fluctuational damping $\overline{\Omega^2_\|}\tau_*$ ($\overline{\Omega^2_\perp}\tau_*$), we obtain

$$4\pi\hat{\sigma}_{kl}(\mathbf{k},w)=iw\left(\frac{\omega_e^2}{w^2-\Omega_e^2}+\frac{\omega_i^2}{w^2-\Omega_i^2}\right)\delta_{kl}$$
$$+\left(\frac{\omega_e^2\Omega_e^2}{w^2-\Omega_e^2}+\frac{\omega_i^2\Omega_i^2}{w^2-\Omega_i^2}\right)\frac{l_kl_l}{iw}$$
$$+\left(\frac{\omega_e^2\Omega_e}{w^2-\Omega_e^2}-\frac{\omega_i^2\Omega_i}{w^2-\Omega_i^2}\right)\varepsilon_{kml}l_m$$
$$+\left(\frac{\omega_e^2\Omega_eh_{\perp e}}{w(w^2-\Omega_e^2)}-\frac{\omega_i^2\Omega_ih_{\perp i}}{w(w^2-\Omega_i^2)}\right)(l_mk_m\delta_{kl}-l_lk_k) \quad (31)$$
$$-i\left(\frac{\omega_e^2h_{\|e}}{w^2}+\frac{\omega_i^2h_{\|i}}{w^2}\right)l_kl_m\varepsilon_{mnl}k_n$$
$$-i\left(\frac{\omega_e^2h_{\perp e}}{w^2-\Omega_e^2}+\frac{\omega_i^2h_{\perp i}}{w^2-\Omega_i^2}\right)(\varepsilon_{kml}k_m-l_kl_m\varepsilon_{lmn}k_n).$$

[2] The characteristic scale of the fluctuational magnetic helicity is known for solar-wind turbulence [24], where it lies within the range 0.004–0.02 AU (~6×10^8–3×10^9 m).

Hence, the permittivity tensor is

$$\varepsilon_{ij} = \delta_{ij} + \frac{4\pi i}{w}\hat{\sigma}_{ij},$$

$$\hat{\varepsilon} = \begin{pmatrix} \varepsilon_\perp + i\chi_0 k_z & ig + \chi_\perp k_z & -i\chi_0 k_x - \chi_\perp k_y \\ -ig - \chi_\perp k_z & \varepsilon_\perp + i\chi_0 k_z & -i\chi_0 k_y + \chi_\perp k_x \\ \chi_\parallel k_y & -\chi_\parallel k_x & \varepsilon_\parallel \end{pmatrix}, \quad (32)$$

where

$$\varepsilon_\perp = 1 - \frac{\omega_e^2}{w^2 - \Omega_e^2} - \frac{\omega_i^2}{w^2 - \Omega_i^2},$$

$$\varepsilon_\parallel = 1 - \frac{\omega_e^2}{w^2} - \frac{\omega_i^2}{w^2}, \quad (33)$$

$$g = \frac{\omega_e^2\Omega_e}{w(\Omega_e^2 - w^2)} - \frac{\omega_i^2\Omega_i}{w(\Omega_i^2 - w^2)},$$

$$\chi_0 = \frac{\omega_e^2\Omega_e}{w^2(w^2 - \Omega_e^2)}h_{\perp e} - \frac{\omega_i^2\Omega_i}{w^2(w^2 - \Omega_i^2)}h_{\perp i}, \quad (34)$$

$$\chi_\perp = \frac{h_{\perp e}}{w}\frac{\omega_e^2}{w^2 - \Omega_e^2} + \frac{h_{\perp i}}{w}\frac{\omega_i^2}{w^2 - \Omega_i^2},$$

$$\chi_\parallel = \frac{h_{\parallel e}}{w}\frac{\omega_e^2}{w^2 - \Omega_e^2} + \frac{h_{\parallel i}}{w}\frac{\omega_i^2}{w^2 - \Omega_i^2}. \quad (35)$$

As we see from (32), fluctuational helicity gives rise to additional gyrotropic terms in the permittivity tensor. To elucidate their role, let us analyze the dispersion relation for electromagnetic waves.

4. THE DISPERSION RELATION

Denote the angle between the vectors **n** and $\mathbf{B}_0$ by θ. The dispersion relation for the complex refractive index $\mathbf{n} = c\mathbf{k}/w$ is defined as [11]

$$\det\left\| n^2\delta_{ij} - n_i n_j - \hat{\varepsilon}_{ij}\right\| = 0. \quad (36)$$

We also assume that

$$\mathbf{n} = (n\sin\theta, 0, n\cos\theta).$$

The dispersion relation is then

$$(g^2 + (n^2 - \varepsilon_\perp)\varepsilon_\perp)\varepsilon_\parallel - n^2\varepsilon_\parallel(n^2 - \varepsilon_\perp)\cos^2\theta$$
$$- n^2(g^2 + (n^2 - \varepsilon_\perp)\varepsilon_\perp)\sin^2\theta$$
$$- n^2\varepsilon_\parallel w^2\frac{-\chi_0^2 + \chi_\perp^2}{c^2}\cos^2\theta$$
$$- n^2 w^2\frac{(g\chi_0 + \varepsilon_\perp\chi_\perp)\chi_\parallel}{c^2}\sin^2\theta \quad (37)$$
$$+ inw\cos\theta\varepsilon_\parallel\frac{n^2\chi_0 - 2\varepsilon_\perp\chi_0 - 2g\chi_\perp}{c}$$
$$+ in^3 w\cos\theta\left(\frac{\varepsilon_\parallel\chi_0}{c}\cos^2\theta + \frac{\varepsilon_\perp\chi_0 + g\chi_\perp - g\chi_\parallel}{c}\sin^2\theta\right) = 0.$$

Let us consider waves that propagate along the magnetic field, θ = 0. In this case, the dispersion relation (37) has the solutions

$$n_{1,2} = \frac{1}{2}\left\{ iw\frac{\chi_0 + \chi_\perp}{c} \pm \left(4(\varepsilon_\perp - g) - w^2\frac{(\chi_0 + \chi_\perp)^2}{c^2}\right)^{1/2}\right\},$$
$$n_{3,4} = \frac{1}{2}\left\{ iw\frac{\chi_0 - \chi_\perp}{c} \pm \left(4(\varepsilon_\perp + g) - w^2\frac{(\chi_0 - \chi_\perp)^2}{c^2}\right)^{1/2}\right\}. \quad (38)$$

Given that

$$\varepsilon_\perp \mp g = 1 - \frac{\omega_e^2}{w(w \pm \Omega_e)} - \frac{\omega_i^2}{w(w \mp \Omega_i)},$$

$$\chi_0 \pm \chi_\perp = \pm\alpha_{\perp e}\frac{\omega_e^2\Omega_e}{w^2(w \mp \Omega_e)\Omega_{\kappa\perp}}c \pm \alpha_{\perp i}\frac{\omega_i^2\Omega_i}{w^2(w \pm \Omega_i)\Omega_{\kappa\perp}}c, \quad (39)$$

the wave vector is

$$ck = i\left(\pm\frac{\alpha_{\perp e}}{2\Omega_{\kappa\perp}}\frac{\omega_e^2\Omega_e}{w \mp \Omega_e} \pm \frac{\alpha_{\perp i}}{2\Omega_{\kappa\perp}}\frac{\omega_i^2\Omega_i}{w \pm \Omega_i}\right)$$
$$\pm\left\{ w^2\left(1 - \frac{\omega_e^2}{w(w \pm \Omega_e)} - \frac{\omega_i^2}{w(w \mp \Omega_i)}\right)\right.$$
$$\left. - \left(\pm\frac{\alpha_{\perp e}}{2\Omega_{\kappa\perp}}\frac{\omega_e^2\Omega_e}{w \mp \Omega_e} \pm \frac{\alpha_{\perp i}}{2\Omega_{\kappa\perp}}\frac{\omega_i^2\Omega_i}{w \pm \Omega_i}\right)^2\right\}^{1/2},$$

whence the equation for frequency is

$$w^2 - w\left(\frac{\omega_e^2}{w \pm \Omega_e} + \frac{\omega_i^2}{w \mp \Omega_i}\right)$$
$$\pm i\frac{ck}{\Omega_{k\perp}}\left(\pm\alpha_{\perp e}\frac{\omega_e^2\Omega_e}{w \mp \Omega_e} \pm \alpha_{\perp i}\frac{\omega_i^2\Omega_i}{w \pm \Omega_i}\right) = c^2k^2. \quad (40)$$

At low frequencies with $\omega_e^2/\Omega_e^2 \ll \omega_i^2/\Omega_i^2$ and $\omega_e^2 \gg \omega_i^2$, the square of the frequency is

$$w^2 = \frac{v_A^2 k^2}{1 + (1 + \alpha_{\perp i}) v_A^2/c^2}\left(1 \pm i\alpha_{\perp e} k\lambda_{\kappa\perp}\frac{\omega_e^2}{c^2 k^2}\right), \quad (41)$$

where $v_A^2 = B_0^2/4\pi nM$.

At low values of $\alpha_{\perp e} k\lambda_{\kappa\perp}\omega_e^2/c^2k^2$ (for small scales),

$$w = \frac{v_A k}{(1 + (1 + \alpha_{\perp i}) v_A^2/c^2)^{1/2}} \pm i\frac{v_A/c}{(1 + (1 + \alpha_{\perp i}) v_A^2/c^2)^{1/2}}\frac{\alpha_{\perp e}\lambda_{\kappa\perp}\omega_e^2}{2c}. \quad (42)$$

In this case, the coefficient of the complex refractive index does not depend on the wave vector. In contrast, at high values of $\alpha_{\perp e} k\lambda_{\kappa\perp}\omega_e^2/c^2k^2$ (for large scales),

$$w = \frac{v_A/c}{(1 + (1 + \alpha_{\perp i}) v_A^2/c^2)^{1/2}} \times\left(\frac{\alpha_{\perp e} k\lambda_{\kappa\perp}}{2}\right)^{1/2}\omega_e(1 \pm i). \quad (43)$$

As we see, in the presence of magnetic fluctuation helicity, there is an instability and the amplitude of the electromagnetic waves propagating in a plasma increases. This demonstrates the nonequilibrium existence of reflectional symmetry breaking at the level of fluctuations. Thus, for example, helicity also leads to an instability, an inverse energy cascade, in magnetohydrodynamics [2]. Unstable waves have nonzero helicity, i.e., a vortex component of the electric field. The motion of charged particles in a magnetic field with fluctuational helicity is equivalent to the motion in random helical magnetic fields with preferred helix orientation. The resonance condition during the motion of particles in a helical magnetic field is satisfied for the particles that move in a direction opposite to the field ($\mathbf{B} \cdot \mathbf{v} < 0$) [25]. After averaging, this resonance condition will correspond to the following: when the helicities of the perturbations and fluctuations have opposite signs, the perturbations will give up energy to particles of the medium; in contrast, when the helicities of the perturbations and fluctuations have the same signs, the field will be amplified—taking away energy from particles of the medium. Indeed, the helicity of growing waves coincides in sign with the small-scale fluctuational helicity. In the opposite case, the perturbation is damped. Note also that the dispersion of the propagating waves changes as well. For large scales, $w \sim k^{1/2}$, the dispersion law is similar to that of gravity waves in deep water whose phase velocity increases with scale. Such long waves can be revealed in the spectrum of geoelectromagnetic perturbations. Note that the fast large-scale electric perturbations in the E region of the ionosphere that accompany such catastrophic events as magnetic storms and substorms, earthquakes, and man-made explosions are, to all appearances, of a vortex nature [26].

Let us consider the range of helicon frequencies: $\Omega_i \ll w \ll \Omega_e$, $\omega_e^2 \gg w\Omega_e$. In this case, the frequency can be expressed as

$$w = \pm\Omega_e\frac{c^2k^2}{\omega_e^2} \mp i\alpha_{\perp e} k\lambda_{\kappa\perp}\Omega_e. \quad (44)$$

The wave propagation is also accompanied by an instability with the growth rate $\alpha_{\perp e} k\lambda_{\kappa\perp}\Omega_e$.

Retaining the quadratic terms in the expansion of the Lorentz force $\left\langle \hat{\tilde{\mathbf{v}}}(\mathbf{q}, s) \times \hat{\tilde{\mathbf{B}}}(\mathbf{k} - \mathbf{q}, w - s)\right\rangle$ in terms of large scales ($k \ll q$) yields a lower limit for such instability [10], and the perturbations are damped at $k > k_{\text{crit}}$.

Let us consider the waves that propagate perpendicular to the magnetic field, $\theta = \pi/2$. In this case, the square of the complex refractive index is

$$n_1^2 = \frac{1}{2\varepsilon_\perp}\{\varepsilon_\perp(\varepsilon_\perp + \varepsilon_\parallel) - g^2 - \kappa + [(\varepsilon_\perp(\varepsilon_\perp - \varepsilon_\parallel) - g^2)^2 - 2(\varepsilon_\perp(\varepsilon_\perp + \varepsilon_\parallel) - g^2)\kappa + \kappa^2]^{1/2}\},$$

$$n_2^2 = \frac{1}{2\varepsilon_\perp}\{\varepsilon_\perp(\varepsilon_\perp + \varepsilon_\parallel) - g^2 - \kappa - [(\varepsilon_\perp(\varepsilon_\perp - \varepsilon_\parallel) - g^2)^2 - 2(\varepsilon_\perp(\varepsilon_\perp + \varepsilon_\parallel) - g^2)\kappa + \kappa^2]^{1/2}\}, \quad (45)$$

$$\kappa = \frac{w^2(g\chi_0 + \varepsilon_\perp\chi_\perp)\chi_\parallel}{c^2}.$$

In the absence of helicity, the first and second expressions in (45) would correspond to the extraordinary and ordinary waves, respectively. As we see, their propagation conditions change, and elliptical polarization attributable to helicity appears in both types of waves.

5. THE OHM LAW FOR LOW FREQUENCIES

Let us consider the case of low frequencies where $w \ll \overline{\Omega_{\parallel_{e(i)}}^2}\tau_*$ ($\overline{\Omega_{\perp_{e(i)}}^2}\tau_*$) and take into account the collision frequency $\nu = 1/\tau_c \gg w$ ($1/\tau_c'$ for ions). To simplify

our calculations, we also assume a weak anisotropy of the fluctuations, $\overline{\Omega^2_{\perp e(i)}} \approx \overline{\Omega^2_{\parallel e(i)}}$ and $\alpha_{\perp e(i)}\lambda_{\kappa\perp} \approx \alpha_{\parallel e(i)}\lambda_{\kappa\parallel} = \alpha_{e(i)}\lambda_\kappa$. The conductivity tensor (28) in this limit will then appear as

$$
\begin{aligned}
4\pi\hat{\sigma}_{kl}(\mathbf{k}, w) &= \left(\frac{\omega_e^2\tau_e}{1+\Omega_e^2\tau_e^2} + \frac{\omega_i^2\tau_i}{1+\Omega_i^2\tau_i^2}\right)\delta_{kl} \\
&+ \left(\frac{\omega_e^2\Omega_e^2\tau_e^3}{1+\Omega_e^2\tau_e^2} + \frac{\omega_i^2\Omega_i^2\tau_i^3}{1+\Omega_i^2\tau_i^2}\right)l_k l_l \\
&- \left(\frac{\omega_e^2\Omega_e\tau_e^2}{1+\Omega_e^2\tau_e^2} - \frac{\omega_i^2\Omega_i\tau_i^2}{1+\Omega_i^2\tau_i^2}\right)\varepsilon_{kml}l_m \\
&+ i\left(\alpha_e\frac{\omega_e^2\Omega_e\tau_e^2}{1+\Omega_e^2\tau_e^2}\left(1+\frac{\tau_e}{\tau_c}\right) - \alpha_i\frac{\omega_i^2\Omega_i\tau_i^2}{1+\Omega_i^2\tau_i^2}\left(1+\frac{\tau_i}{\tau_c'}\right)\right) \\
&\times \lambda_\kappa(l_m k_m\delta_{kl} - l_l k_k) \\
&+ i(\alpha_e\omega_e^2(\tau_c+\tau_e) + \alpha_i\omega_i^2(\tau_c'+\tau_i))\lambda_\kappa l_k l_m\varepsilon_{mnl}k_n \\
&+ i\left(\alpha_e\frac{\omega_e^2(\tau_c+\tau_e)}{1+\Omega_e^2\tau_e^2} + \alpha_i\frac{\omega_i^2(\tau_c'+\tau_i)}{1+\Omega_i^2\tau_i^2}\right) \\
&\times \lambda_\kappa(\varepsilon_{kml}k_m - l_k l_m\varepsilon_{lmn}k_n).
\end{aligned}
\tag{46}
$$

Here, we introduced the following characteristic time scales:

$$
\tau_e = \frac{\tau_c}{1+\overline{\Omega^2_{\parallel e}}\tau_*\tau_c}, \quad \tau_i = \frac{\tau_c'}{1+\overline{\Omega^2_{\parallel i}}\tau_*'\tau_c}. \tag{47}
$$

Having defined the conductivities

$$
\sigma_{0e(i)} = \frac{\omega^2_{e(i)}\tau_{e(i)}}{4\pi}, \quad \sigma_{e(i)\perp} = \frac{\sigma_{0e(i)}}{1+\Omega^2_{e(i)}\tau^2_{e(i)}},
$$

$$
\sigma_{e(i)\parallel} = \frac{\sigma_{0e(i)}\Omega^2_{e(i)}\tau^2_{e(i)}}{1+\Omega^2_{e(i)}\tau^2_{e(i)}},
$$

the current after applying the inverse Fourier transform can be written as

$$
\begin{aligned}
\mathbf{j} &= (\sigma_{e\perp}+\sigma_{i\perp})\langle\mathbf{E}\rangle + (\sigma_{e\parallel}+\sigma_{i\parallel})\mathbf{l}(\mathbf{l}\cdot\langle\mathbf{E}\rangle) \\
&- (\sigma_{e\perp}\Omega_e\tau_e - \sigma_{i\perp}\Omega_i\tau_i)[\mathbf{l}\times\langle\mathbf{E}\rangle] \\
&+ \left(\alpha_e\sigma_{e\perp}\Omega_e\tau_e\left(1+\frac{\tau_e}{\tau_c}\right) - \alpha_i\sigma_{ie\perp}\Omega_i\tau_i\left(1+\frac{\tau_i}{\tau_c'}\right)\right) \\
&\times \lambda_\kappa((\mathbf{l}\cdot\nabla)\langle\mathbf{E}\rangle - \nabla(\mathbf{l}\cdot\langle\mathbf{E}\rangle)) \\
&+ \left(\alpha_e(\sigma_{0e}-\sigma_{e\perp})\left(1+\frac{\tau_c}{\tau_e}\right)\right. \\
&\left. + \alpha_i(\sigma_{0i}-\sigma_{i\perp})\left(1+\frac{\tau_c'}{\tau_i}\right)\right)\lambda_\kappa\mathbf{l}(\mathbf{l}\cdot\mathrm{curl}\langle\mathbf{E}\rangle) \\
&+ \left(\alpha_e\sigma_{e\perp}\left(1+\frac{\tau_c}{\tau_e}\right) + \alpha_i\sigma_{i\perp}\left(1+\frac{\tau_c'}{\tau_i}\right)\right)\lambda_\kappa\mathrm{curl}\langle\mathbf{E}\rangle.
\end{aligned}
\tag{48}
$$

The influence of fluctuations and external magnetic field primarily causes the conductivity to decrease, while the presence of helicity leads to an additional dependence of the current on the vortex component of the electric field.

Let us consider the mean magnetic field in a medium with the Ohm law (48). We will disregard the ion component (the EMHD approximation). Neglecting the displacement current, we obtain the following expression for the growth rate γ of wave field perturbations of the form

$$
\begin{aligned}
\langle\mathbf{E}\rangle &= \exp(\gamma t)(E_x(z), E_y(z), 0), \\
\langle\mathbf{B}\rangle &= \exp(\gamma t)(B_x(z), B_y(z), 0):
\end{aligned}
$$

$$
\gamma = -\frac{c^2k^2}{4\pi\sigma_{e\perp}}\frac{\tau_c + i\alpha_e k\lambda\zeta\Omega_e\tau_e}{((1-i\Omega_e\tau_e) - \alpha_e k\lambda\zeta(\tau_c/\tau_e - i\Omega_e\tau_e))((1+i\Omega_e\tau_e) + \alpha_e k\lambda\zeta(\tau_c/\tau_e + i\Omega_e\tau_e))}, \tag{49}
$$

where $\zeta = 1 + \tau_e/\tau_c$.

For the wave vectors

$$
\alpha_e|k\lambda| > \frac{(1+\Omega_e^2\tau_e^2)^{1/2}}{(1+\tau_e/\tau_c)(\tau_c^2/\tau_e^2 + (2\tau_c/\tau_e - 1)\Omega_e^2\tau_e^2)^{1/2}}
$$

the perturbations grow. For intense magnetic fluctuations, $\tau_e \approx 1/\overline{\Omega^2_\parallel}\tau_* \ll \tau_c$, and we obtain for the threshold wave number

$$
\alpha_e|\kappa\lambda| \gtrsim \frac{1}{\overline{\Omega^2_\parallel}\tau_*\tau_c}\left(1 + \frac{\Omega_e^2}{2(\overline{\Omega^2_\parallel}\tau_*)^2}\right).
$$

In the collisionless limit $\tau_c \longrightarrow 0$, the threshold wave number is

$$\alpha_e|k\lambda| > \frac{1}{2};$$

i.e., the threshold instability scale also increases with fluctuation amplitude (parameter α_e). On this threshold scale, the waves with the following frequency propagate at $\tau_e \ll \tau_c$:

$$w = \frac{c^2}{8\pi\sigma_{0e}}\frac{1}{\alpha_e^2\lambda^2\overline{\Omega_\parallel^2}\tau_*\tau_c}\frac{1+\Omega_e^2\tau_c^2}{\Omega_e\tau_c}\left(1+\frac{\Omega_e^2}{2(\overline{\Omega_\parallel^2}\tau_*\tau_c)^2}\right).$$

Retaining the quadratic terms in the permittivity tensor gives rise to dissipative terms of the form $-\sigma_*\Delta\langle\mathbf{E}\rangle + \sigma_*'\nabla\mathrm{div}\langle\mathbf{E}\rangle$ [10] in the Ohm law. Their influence restricts the instability region, and the field perturbations are damped on small scales.

6. FINITE CORRELATION TIMES

Consider the effects of finite correlation times for high frequencies, $w\tau \gg 1$, with the anisotropy effects disregarded. These also include the case of long correlation times. In this limit, the effective Lorentz force is

$$\frac{e}{mc}\int\langle\hat{\mathbf{v}}(\mathbf{q},s)\times\hat{\mathbf{B}}(\mathbf{k}-\mathbf{q},w-s)\rangle\,d\mathbf{q}\,ds$$

$$= -\frac{4}{3}iw\tau\left(\frac{e}{mc}\right)^2\hat{\mathscr{E}}\tau_*\langle\hat{\mathbf{v}}(\mathbf{k},w)\rangle$$

$$-\frac{2\tau}{3}\frac{e}{m}\left(\frac{e}{mc}\right)^2H_{0\perp}\tau_*i[\mathbf{k}\times\langle\hat{\mathbf{E}}(\mathbf{k},w)\rangle]$$

$$+\frac{2}{3}\left(\frac{e}{mc}\right)^2E_0\tau_*\tau[\Omega_\mathbf{e}\times\langle\hat{\mathbf{v}}(\mathbf{k},w)\rangle]$$

$$-\frac{2}{3}i\tau w\left(\frac{e}{mc}\right)^2H_0\tau_*\Omega_\mathbf{e}\delta(\mathbf{k})\delta(w).$$

The gyrotropic fluctuational acceleration will be replaced with oscillations. The frequencies will acquire a negative shift, and the permittivity tensor will take the form

$$\hat{\varepsilon} = \begin{pmatrix} \varepsilon_\perp - w\tau\chi_0 k_z & ig + iw\tau\chi_\perp k_z & w\tau\chi_0 k_x - iw\tau\chi_\perp k_y \\ -ig - iw\tau\chi_\perp k_z & \varepsilon_\perp - w\tau\chi_0 k_z & w\tau\chi_0 k_y + iw\tau\chi_\perp k_x \\ iw\tau\chi_\parallel k_y & -iw\tau\chi_\parallel k_x & \varepsilon_\parallel \end{pmatrix}. \tag{50}$$

Consider the waves propagating along the magnetic field, $\theta = 0$. In this case, the dispersion relation (37) has the solutions

$$n_{1,2} = (\varepsilon_\perp + g)\left(1 \pm \frac{2w\tau(\chi_0-\chi_\perp)}{(4c^2(\varepsilon_\perp+g)^2+(\chi_0-\chi_\perp)^2w^4\tau^2)^{1/2} \mp (\chi_0-\chi_\perp)cw^2\tau}\right),$$
$$n_{3,4} = (\varepsilon_\perp - g)\left(1 \pm \frac{2w^2\tau(\chi_0+\chi_\perp)}{(4c^2(\varepsilon_\perp-g)^2+(\chi_0-\chi_\perp)^2w^4\tau^2)^{1/2} \mp (\chi_0+\chi_\perp)cw^2\tau}\right). \tag{51}$$

Assuming the helical additions to be small, we can write

$$n_{1,2} = \varepsilon_\perp + g \pm \frac{w^2\tau(\chi_0-\chi_\perp)}{c}$$
$$= 1 - \frac{\omega_e^2}{w(w-\Omega_e)} - \frac{\omega_i^2}{w(w+\Omega_i)} \tag{52}$$
$$\mp\left(\alpha_{\perp e}\frac{\omega_e^2\Omega_e\tau}{(w+\Omega_e)\Omega_{\kappa\perp}} + \alpha_{\perp i}\frac{\omega_i^2\Omega_i\tau}{(w-\Omega_i)\Omega_{\kappa\perp}}\right),$$

$$n_{3,4} = \varepsilon_\perp - g \pm \frac{w^2\tau(\chi_0+\chi_\perp)}{c}$$
$$= 1 - \frac{\omega_e^2}{w(w+\Omega_e)} - \frac{\omega_i^2}{w(w-\Omega_i)} \tag{53}$$
$$\pm\left(\alpha_{\perp e}\frac{\omega_e^2\Omega_e\tau}{(w-\Omega_e)\Omega_{\kappa\perp}} + \alpha_{\perp i}\frac{\omega_i^2\Omega_i\tau}{(w+\Omega_i)\Omega_{\kappa\perp}}\right).$$

It is easy to see that an additional rotation of the polarization plane appears here.

For the waves propagating perpendicular to the magnetic field, $\theta = \pi/2$, we obtain the following solutions:

$$n_1^2 = \frac{1}{2\varepsilon_\perp}\{\varepsilon_\perp(\varepsilon_\perp + \varepsilon_\parallel) - g^2 + \kappa'$$
$$+ [(\varepsilon_\perp(\varepsilon_\perp - \varepsilon_\parallel) - g^2)^2$$
$$- 2(\varepsilon_\perp(\varepsilon_\perp + \varepsilon_\parallel) - g^2)\kappa' + \kappa'^2]^{1/2}\},$$
$$n_2^2 = \frac{1}{2\varepsilon_\perp}\{\varepsilon_\perp(\varepsilon_\perp + \varepsilon_\parallel) - g^2 + \kappa' \quad (54)$$
$$- [(\varepsilon_\perp(\varepsilon_\perp - \varepsilon_\parallel) - g^2)^2$$
$$- 2(\varepsilon_\perp(\varepsilon_\perp + \varepsilon_\parallel) - g^2)\kappa' + \kappa'^2]^{1/2}\},$$
$$\kappa' = \frac{w^4\tau^2(g\chi_0 + \varepsilon_\perp\chi_\perp)\chi_\parallel}{c^2}.$$

As in the approximation of a δ-correlated random process (45) considered above, the propagation conditions change, and elliptical polarization attributable to helicity appears in both the ordinary and extraordinary waves. Note that in the case of infinite correlation times or high frequencies (frozen fluctuations), the properties of a plasma medium with magnetic helicity are similar to those of chiral and bianisotropic media [14, 15].

In actual systems, $w\tau$ is finite and the effects of both instability considered in Section 4 and the appearance of additional wave modes must simultaneously manifest themselves.

7. CONCLUSIONS

The influence of magnetic fluctuations on the motion of the particles of a cold magnetoactive plasma primarily reduces to the appearance of an effective fluctuational collision frequency determined by the statistical parameters and to a decrease in conductivity. Reflectional symmetry breaking—nonzero mean magnetic helicity of the fluctuations—leads to a change in the dispersion of the propagating waves and the appearance of additional modes. The waves can be unstable, reflecting both the nonequilibrium nature of the turbulent magnetic helicity and the peculiarities of the particle motion in random helical magnetic fields. The instability growth rate is proportional to the helicity of the fluctuational magnetic field and the amplitude of the large scale uniform magnetic field. Allowance for the finite correlation times and for the additional fluctuational quadratic dispersion effects restricts the action of this instability. In contrast to the turbulent dynamo effects considered in the MHD and EMHD approximations [2], here there is a natural restriction on the instability region on large scales determined by the relationship between the fluctuational helicity and energy and the large-scale magnetic field. The plasma acquires properties similar to those observed in chiral and bianisotropic media [14, 15]. Consequently, it can have properties characteristic of these media, such as anomalous absorption [27, 28] and additional wave conversion effects [29, 30]. In contrast to the artificial external origin of the chirality in chiral media, this property is natural in a turbulent magnetoactive plasma with helicity. The deviations in the rotation of the polarization plane attributable to fluctuational helicity can serve as a tool for diagnosing it. The results were obtained in the approximation of isolated particles, the advantages of which and disadvantages are well known. It is easy to see that the above effects are preserved when the thermal and collisional effects are taken into account and can be obtained in terms of the kinetic approach.

ACKNOWLEDGMENTS

I am grateful to S.N. Artekh and N.S. Erokhin for helpful discussions. This work was supported by the Foundation for Support of Russian Science.

REFERENCES

1. J. B. Taylor, Phys. Rev. Lett. **33**, 1139 (1974).
2. S. I. Vaĭnshteĭn, Ya. B. Zel'dovich, and A. A. Ruzmaĭkin, *The Turbulent Dynamo in Astrophysics* (Nauka, Moscow, 1980) [in Russian].
3. K. Hasselmann and G. Wibberenz, Z. Geophys. **34**, 353 (1968).
4. M. L. Goldstein and W. H. Matthaeus, in *Proceedings of 17th International Cosmic Ray Conference* (1981), Vol. 3, p. 294.
5. L. L. Kichatinov, Pis'ma Zh. Éksp. Teor. Fiz. **37**, 43 (1983) [JETP Lett. **37**, 51 (1983)].
6. Yu. I. Fedorov, V. E. Katz, L. L. Kichatinov, and M. Stehlic, Astron. Astrophys. **260**, 499 (1992).
7. R. R. Mett and J. A. Tataronis, Phys. Rev. Lett. **63**, 1380 (1989).
8. J. B. Taylor, Phys. Rev. Lett. **63**, 1384 (1989).
9. A. V. Chechkin, V. V. Yanovsky, and A. V. Tur, Phys. Plasmas **1**, 2566 (1994).
10. O. G. Chkhetiani and S. S. Moiseev, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 268 (1999) [JETP Lett. **70**, 277 (1999)].
11. V. L. Ginzburg and A. A. Rukhadze, *Waves in Magnetized Plasmas* (Nauka, Moscow, 1975) [in Russian].
12. S. I. Vaihnstein, Phys. Rev. Lett. **80**, 4879 (1998).
13. A. F. Aleksandrov and A. A. Rukhadze, *Lectures on Electrodynamics of Plasma-Like Media* (Mosk. Gos. Univ., Moscow, 1999) [in Russian].
14. D. L. Jaggard, A. R. Mickelson, and C. H. Papas, Appl. Phys. **18**, 211 (1979).
15. B. Z. Katsenelenbaum, E. N. Korshunova, A. N. Sizov, and A. D. Shatrov, Usp. Fiz. Nauk **167**, 1201 (1997) [Phys. Usp. **40**, 1149 (1997)].

16. E. A. Novikov, Zh. Éksp. Teor. Fiz. **47**, 1919 (1964) [Sov. Phys. JETP **20**, 1290 (1964)].
17. V. I. Klyatskin and V. I. Tatarskiĭ, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **15**, 1433 (1972).
18. S. I. Vaĭnshteĭn, Zh. Éksp. Teor. Fiz. **58**, 153 (1970) [Sov. Phys. JETP **31**, 87 (1970)].
19. W. H. Matthaeus and C. Smith, Phys. Rev. A **24**, 2135 (1981).
20. S. Oughton, K.-H. Rädler, and W. H. Matthaeus, Phys. Rev. E **56**, 2875 (1997).
21. F. Krause and K.-H. Radler, *Mean-Field Magnetohydrodynamics and Dynamo Theory* (Akademie, Berlin, 1980; Mir, Moscow, 1984).
22. I. N. Toptygin, *Cosmic Rays in Interplanetary Magnetic Fields* (Nauka, Moscow, 1983; Reidel, Dordrecht, 1985).
23. A. V. Chechkin, D. P. Sorokin, and V. V. Yanovsky, hep-th/9306159.
24. C. W. Smith and J. W. Bieber, AIP Conf. Proc. **382**, 498 (1996).
25. A. I. Morozov and L. S. Solov'ev, Zh. Tekh. Fiz. **30**, 271 (1960) [Sov. Phys. Tech. Phys. **5**, 250 (1960)].
26. A. G. Khantadze, G. D. Aburdzhania, G. V. Dzhandieri, and Kh. Z. Chargaziya, Fiz. Plazmy (Moscow) **30**, 88 (2004) [Plasma Phys. Rep. **30**, 83 (2004)].
27. D. L. Jaggard and N. Engheta, Electron. Lett. **25**, 1060 (1989).
28. V. A. Neganov and O. V. Osipov, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **42**, 870 (1999).
29. N. S. Erokhin and S. S. Moiseev, Preprint No. 1948, IKI AN RAN (Inst. for Space Research, Russian Academy of Sciences, Moscow, 1996).
30. H. T. Torres, P. H. Sakanaka, and N. Reggiani, J. Phys. Soc. Jpn. **67**, 850 (1998).

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 99, No. 2, 2004, pp. 340–342.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 2, 2004, pp. 381–384.

FLUIDS

Convective Boundary Layers in Electrolyte Solutions

A. P. Grigin

Frumkin Institute of Electrochemistry, Russian Academy of Sciences, Moscow, 119071 Russia
e-mail: grigin@ns.elchem.ac.ru

Received January 21, 2004

Abstract—During current passage through an electrolyte solution, ions entrain their hydration shells, which gives rise to the accompanying diffusion flow of water molecules and internal convection analogous to that observed in superfluid He^3 + He^4 solutions. Hydrodynamic conditions are determined under which the total water flux on a solid surface is zero. The convective boundary layer on a vertical plate is theoretically described using the von Karman method. The diffusion–mechanical effect in thin capillaries is evaluated. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

As is known, the motion of an electron in an ionic crystal may be accompanied by the phenomenon of autolocalization of the charged particle in a polarization well. The motion of this quasiparticle—a polaron—involves the motion of the potential well together with the electron. However, since dipoles are fixed at the lattice sites, the potential well of a moving polaron disappears upon its leaving one lattice site and appears again as it passes by another site and so on.

In polar liquids, the motion of charged particles may proceed according to an alternative mechanism. According to this, a polarization well appears at an ion (by analogy with the case of ionic crystals) since the charged particle is surrounded by polarized water molecules forming the hydration shell. The hydration shell is formed due to chemical bonds—quantum exchange interactions between ions and water molecules, obeying the Pauli principle of the wave function symmetry. If the chemical bonds formed between an ion and the surrounding water molecules are sufficiently strong, it is energetically favorable for the charged particle to move together with the hydration shell (retaining the same water molecule in the surrounding), rather than to periodically destruct and reconstruct this polarization well. Thus, in an electrolyte solution, the diffusion flow of ions is accompanied by the flow of bound water molecules. On a metal surface under the conditions of electrolysis, ions pass from solution to the ion lattice, whereas water molecules of their hydration shells accumulate at the surface until the internal convection will arise, representing a hydrodynamic flow of liquid generated at the metal surface. Naturally, the internal convection in electrolyte solutions differs from that in superfluid helium; nevertheless, an important similarity still exists.

The phenomenon of internal convection is well known in the physics of quantum fluids, where this term refers to counterflowing normal and superfluid components arising in quantum solutions in the presence of a temperature gradient. The flux of the superfluid component is directed along the temperature gradient (from the region of lower temperature toward heated fluid) [1]. The two-fluid hydrodynamics (i.e., the hydrodynamics with internal convection) predicts a number of special effects. In particular, convective instability in the quantum He^3 + He^4 solutions arises on heating from below, rather from above as in the normal fluids [2]. On the other hand, the analogs of the two-fluid effects, such as thermomechanical effect, are also observed in normal fluids [3]. We can expect that internal convection caused by the diffusion transfer of water molecules by ions in electrolyte solutions will also give rise to new effects.

2. HYDRODYNAMIC BOUNDARY CONDITIONS

Let us find a relationship between the normal velocity component v_n of a liquid flowing over a metal surface and the flux I of ions to this surface. The number n_{H_2O} of water molecules generated on a unit electrode surface area per unit time obeys the equation

$$n_{H_2O} = Ir, \tag{1}$$

where r is the coordination number (the number of water molecules strongly bound to and moving with an ion). According to this relation, the mass density Q_{H_2O} of the diffusion flux of water on the electrode surface is

$$Q_{H_2O} = \frac{M_{H_2O} rI}{N}, \tag{2}$$

where M_{H_2O} is the gram-molecular weight of water and N is the Avogadro number. Equation (2) describes the mass flow rate of water from the volume of the solution to the cathode surface. Since the metal surface is impermeable to water molecules, the total flux on this surface

must be zero. Therefore, the metal surface must obey the following condition:

$$v_n\rho + Q_{H_2O} = 0, \quad (3)$$

where ρ is the liquid density. Using formula (3), we readily obtain an expression for the velocity of internal convection on the metal surface:

$$v_n = -\frac{M_{H_2O} r}{N\rho} I. \quad (4)$$

Thus, the boundary conditions for the normal and tangential velocity components on the metal surface can be written as

$$v_n|_s = -\alpha I|_s, \quad v_t|_s = 0, \quad (5)$$

where $\alpha = M_{H_2O}\, r/N\rho$.

Note that the boundary conditions (5) have the same form as those for the normal component of superfluid HeII. Therefore, we may expect that electrolyte solutions are featuring the effects similar to the two-fluid effects in superfluid helium.

3. CONVECTIVE BOUNDARY EFFECTS IN ELECTROLYTE SOLUTIONS

Consider a vertical metal plate (electrode) and let ions of the same metal deposit from an electrolyte solution on the electrode surface. As a result of the ion deposition, the concentration of ions near the electrode surface decreases. This leads to a decrease in the electrolyte density and causes the appearance of buoyancy forces, which set the solution in motion and gives rise to the free concentration-induced convection. For sufficiently large Rayleigh numbers ($Ra > 10^4$), the free convection can be considered in the boundary layer approximation.

Placing the origin of the coordinates on the lower end of the metal plate and directing the x axis upward and the y axis perpendicularly to the metal surface, we can write the following system of equations for the convective boundary layer in dimensionless coordinates [4]:

$$\frac{D}{v}(\mathbf{V}\cdot\nabla)V_x = \frac{\partial^2 V_x}{\partial y^2} + \frac{gH^3 c_0}{vD\rho}\frac{\partial\rho}{\partial c}C, \quad (6)$$

$$\frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y} = 0, \quad (7)$$

$$-\frac{\partial^2 C}{\partial y^2} + V_x\frac{\partial C}{\partial x} + V_y\frac{\partial C}{\partial y} = 0, \quad (8)$$

where D is the ion diffusion coefficient. Dimensionless coordinates are introduced using the electrode height H as the unit of length, the D/H ratio as the unit of velocity, and the Dc_0/H as the unit of particle flux density.

System of equations (6)–(8) was solved using the method of von Karman [5]. According to this method, the boundary conditions for the particle velocity and concentration are formulated on a finite interval $(0, \delta)$ (where $\delta = \delta(x)$ is the boundary layer thickness), rather than on the semi-infinite interval $(0, \infty)$). In the von Karman (boundary layer) approximation, the boundary conditions corresponding to the complete absorption of ions on the metal surface have the following form:

$$V_x|_{y=0} = 0, \quad V_x|_{y=\delta} = 0, \quad \left.\frac{\partial V_x}{\partial y}\right|_{y=\delta} = 0, \quad (9)$$

$$V_y|_{y=0} = -\gamma I, \quad (10)$$

$$C|_{y=0} = 0, \quad C|_{y=\delta} = 1, \quad \left.\frac{\partial C}{\partial y}\right|_{y=\delta} = 0, \quad (11)$$

where

$$\gamma = \frac{M_{H_2O} r}{N\rho} c_0.$$

According to the von Karman method, we pass from differential equations (6) and (8) to the corresponding integral equations. Integrating Eqs. (6) and (8) with respect to y from $y = 0$ to $y = \delta$ and taking into account the boundary conditions (9) and (11), we arrive at the system of equations

$$0 = -\left.\frac{\partial V_x}{\partial y}\right|_{y=0} + \mathrm{Ra}\int_0^\delta (C-1)\,dy, \quad (12)$$

$$\left.\frac{\partial C}{\partial y}\right|_{y=0} + \int_0^\delta \left[V_x\frac{\partial(C-1)}{\partial x} + V_y\frac{\partial(C-1)}{\partial y}\right]dy = 0. \quad (13)$$

Equation (12) is written in the zero approximation with respect to the small parameter D/v, and the nonlinear term is omitted. Calculating the integral in Eq. (13) by parts and using the boundary condition (10), Eq. (13) can be transformed to

$$\left.\frac{\partial C}{\partial y}\right|_{y=0} - \gamma I + \frac{d}{dx}\int_0^\delta V_x(C-1)\,dy = 0. \quad (14)$$

In the von Karman approach adopted, the unknown quantities—the solution velocity and the ion concentration—are found in the form of polynomials satisfying boundary conditions (9) and (11). The simplest polynomials of this type are as follows:

$$V_x = U\frac{y}{\delta}\left(1-\frac{y}{\delta}\right)^2, \quad (15)$$

$$C - 1 = -\left(1-\frac{y}{\delta}\right)^2, \quad (16)$$

where U and δ are functions of the variable x.

Using expression (16), the ion flux to the metal surface can be written as

$$I = -\frac{\partial C}{\partial y}\bigg|_{y=0} = -\frac{2}{\delta}. \quad (17)$$

Using this relation, we obtain the following system of equations for the functions U and δ:

$$-\frac{U}{\delta} + \frac{\mathrm{Ra}\delta}{3} = 0, \quad (18)$$

$$\frac{2+\gamma}{\delta} - \frac{d}{dx}\left(\frac{U\delta}{30}\right) = 0. \quad (19)$$

Integrating this system with the boundary conditions $\delta|_{x=0} = 0$ and using relation (17), we obtain an expression for the current density,

$$I = -2\left[\frac{\mathrm{Ra}}{(1+\gamma)240}\right]^{1/4} x^{1/4}, \quad (20)$$

where

$$\mathrm{Ra} = \frac{gH^3 c_0}{\nu D\rho}\frac{\partial\rho}{\partial c}$$

is the Rayleigh concentration number.

Finally, integrating Eq. (20) over the entire surface plate, was can determine the Sherwood number as

$$\mathrm{Sh} = \frac{8}{3}\left[\frac{\mathrm{Ra}}{240(1+\gamma)}\right]^{1/4} x^{1/4}. \quad (21)$$

As can be seen, the internal convection in electrolyte solutions leads to an increase in thickness of the convective boundary layer and a decrease in the Sherwood number.

4. CONCLUSIONS

It was established that the diffusion flow of ions, which are strongly bound to their hydration shells, in electrolyte solutions is accompanied by the diffusion flow of water. If the ions reaching an electrode pass from solution into metal and release their hydration shells, internal convection appears exhibiting a certain similarity with the internal convection known in superfluid solutions. Since the total flux of water molecules on the metal surface must be zero, the system obeys boundary condition (3), which implies that the solution velocity meets boundary conditions (5). These equations have the same form as the boundary conditions for the normal velocity component of superfluid helium (if the particle flux I is replaced by the flux of the superfluid component). Therefore, we may expect that electrolyte solutions feature the effects analogous to the two-fluid effects in superfluid helium.

For example, let us consider the diffusion-mechanical effect in thin capillaries, which is an analog of the thermomechanical effect in HeII. The difference in the ion concentrations $\Delta c = c_2 - c_1$ between the ends of a thin capillary of length L and radius R gives rise to the diffusion flux

$$I = -\pi r R^2 \frac{D\Delta c}{L}.$$

The ion flow is accompanied by a diffusion flow of water molecules with the mass flux density $Q = \alpha I$. Using the boundary conditions, according to which the total flux of solvent through the capillary cross section is zero, we obtain an expression for the corresponding pressure difference between the ends of the capillary:

$$\Delta p = \alpha\frac{8\eta D\Delta c}{R^2}, \quad (22)$$

where η is the dynamic viscosity. Formula (22) ignores the convective flow of ions and, hence, is valid for

$$\gamma = \frac{M_{\mathrm{H_2O}} r}{N\rho}\Delta c \ll 1.$$

As can be seen from formula (21), the manifestation of internal convection and the two-fluid effects in electrolyte solutions are determined by the dimensionless parameter γ proportional to the concentration of particles (11). The value of this parameter for $c_0 = 6 \times 10^{20}$ cm^{-3} and $r = 10$ is $\gamma = 1.8$; therefore, for not very strongly diluted solutions, the γ value is on the order of unity.

In view of boundary conditions (5), we can ascertain that the hydrodynamics of electrolyte solutions presents an intermediate case between the hydrodynamics of normal fluids and superfluid systems.

REFERENCES

1. I. M. Khalatnikov, *The Theory of Superfluidity* (Nauka, Moscow, 1971) [in Russian].
2. A. P. Grigin, Zh. Éksp. Teor. Fiz. **97**, 1853 (1990) [Sov. Phys. JETP **70**, 1045 (1990)].
3. A. F. Andreev, Zh. Éksp. Teor. Fiz. **59**, 1819 (1970) [Sov. Phys. JETP **32**, 987 (1970)].
4. B. Gebhart, I. Jaluriya, R. Mahajan, and B. Sammakiya, *Buoyancy-Induced Flows and Transport* (Harper and Row, Washington, 1988; Mir, Moscow, 1991), Vol. 1.
5. T. von Karman, Z. Angew. Math. Mech. **1**, 233 (1921).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 2, 2004, pp. 343–351.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 2, 2004, pp. 385–394.

FLUIDS

Polarization of Light Transmitted through a Polymer Film with Nanosized Droplets of Liquid Crystal

V. A. Loiko and A. V. Konkolovich

Stepanov Institute of Physics, National Academy of Sciences of Belarus, Minsk, 220072 Belarus
e-mail: loiko@dragon.bas-net.by

Received March 2, 2004

Abstract—A method is developed for analyzing the state of polarization of a plane wave transmitted through a polymer-dispersed liquid-crystal (PDLC) film with nanosized liquid-crystal (LC) droplets. This method is based on the anisotropic-dipole approximation for describing scattering by a separate droplet and on the Foldy–Twersky approximation for describing propagation of light in a film. Equations are obtained that relate the ellipsometric parameters of coherent (direct) light transmitted through a PDLC film to the order parameters that characterize the morphological and structural properties of the film. Elliptic and circular polarizations and the rotation of the plane of polarization of a wave transmitted through a film are investigated under the normal illumination of the PDLC film by a linearly polarized plane wave. The order parameters of the PDLC film are determined as a function of a control field under the transition from a partially ordered structure of optical axes of LC droplets to a homeotropic structure. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Polymer-dispersed liquid-crystal (PDLC) films are polymer films that contain embedded liquid-crystal (LC) droplets enclosed between two transparent plates with transparent electrodes [1–5]. These films are used in display units and devices for controlling and modulating light beams as diffraction gratings, lenses, polarizers, spectral filters, etc. [3–11]. They are characterized by higher light stability, mechanical strength, flexibility, and resistance to external actions compared with the devices based on homogeneous LC films. The application of PDLC films allows one to enhance the functional capabilities of optical LC elements.

Films with nanosized LC droplets [12–14] represent a new type of PDLC films. These films are characterized by low scattering and high transmission. They can be used for modulating the phase and the polarization of light in telecommunication systems [15]. The development of methods for describing light propagation that allow one to relate the amplitude, phase, and polarization of a transmitted wave to the morphological parameters of a film is an important problem in investigating PDLC films with nanosized LC droplets.

In this paper, we propose a method for analyzing the polarization of a plane wave transmitted through a PDLC film with nanosized droplets of a nematic LC with axially symmetric distribution of molecules. We investigate elliptically and circularly polarized waves and the rotation of the polarization plane of the transmitted wave under the normal illumination of the PDLC film by a linearly polarized plane wave. First, we solve the problem of scattering by a separate LC droplet and then apply the Foldy–Twersky equation in the approximation of independent scatterers to determine the coherent field transmitted through a plane-parallel PDLC film. The specific feature of the approach proposed in this paper is the application of order parameters [16–18] that characterize the orientation ordering of LC molecules inside the droplets and of the optical axes (directors) of the droplets in the film. This approach significantly simplifies solving the problem.

In Section 2, we determine the amplitude scattering matrix averaged over the sizes of LC droplets and the orientations of their optical axes in the anisotropic-dipole approximation [19].

In Section 3, we apply the generalized Foldy–Twersky integral equation [20] for the vector case to derive formulas for the rotation angle of the polarization plane and to derive conditions under which a transmitted wave is circularly polarized.

In Section 4, we present the results of calculations and the analysis of these results for the films in which a transition from a partially ordered structure of directors of nematic droplets to a homeotropic structure occurs. We derive a formula for the order parameters of a film under such an orientation rearrangement of droplet directors as a function of a control field. The results obtained allow one to evaluate a control voltage that admits efficient transformation of polarization.

2. AMPLITUDE SCATTERING MATRIX AND ORDER PARAMETERS OF DROPLETS

Suppose that the amplitude of the field inside a droplet is equal to that of the incident wave and that the variation of the phase of the wave is insignificant. In this

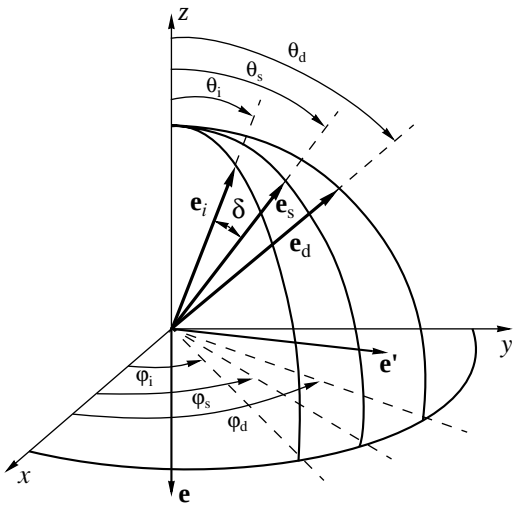


Fig. 1. Geometry of scattering by a separate droplet. Notation is explained in the text.

case, one can apply the Rayleigh approximation [19] to describe the scattering by a separate droplet. Let us represent an LC droplet with a cylindrically symmetric distribution of molecules as an anisotropic dipole. The direction of the induced dipole moment **P** of such a droplet does not coincide with the electric field vector $\mathbf{E}_i$ of the incident wave; i.e.,

$$\mathbf{P} = \varepsilon_p \tilde{\underline{\alpha}} \mathbf{E}_i, \tag{1}$$

where ε_p is the dielectric permittivity of the polymer matrix and $\tilde{\underline{\alpha}}$ is the polarizability tensor of LC droplets in the system of coordinates $\mathbf{ee'e}_i$ fixed to the incident wave (Fig. 1).

In Fig. 1, the unit vector $\mathbf{e}_i$ defines the direction of illumination in the laboratory system of coordinates *xyz*, and $\mathbf{k}_i = k\mathbf{e}_i$ is the wave vector of the incident wave ($k = 2\pi/\lambda_p$, λ_p is the wavelength of the incident wave in the polymer matrix). The unit vector **e** defines the direction of polarization of the incident wave ($\mathbf{E}_i = E_i\mathbf{e}$, E_i is the amplitude of the incident wave). The unit vector **e'** is orthogonal to the polarization plane $\mathbf{e}_i\mathbf{e}$ of the incident wave and is directed along the vector product $\mathbf{e}_i \times \mathbf{e}$. $\mathbf{e}_s$ is the unit vector that defines the direction of scattering ($\mathbf{k}_s = k\mathbf{e}_s$ is the wave vector of the scattered wave). The unit vector $\mathbf{e}_d$ defines the direction of the symmetry axis of the distribution of LC molecules inside a droplet (the optical axis, or the director of the droplet [3, 4, 21]). The angles θ_i, θ_s, and θ_d define the orientations of the vectors $\mathbf{e}_i$, $\mathbf{e}_s$, and $\mathbf{e}_d$ with respect to the *z* axis, and angles φ_i, φ_s, and φ_d define the orientations of the projections of vectors $\mathbf{e}_i$, $\mathbf{e}_s$, and $\mathbf{e}_d$ onto the *xy* plane with respect to the *x* axis. Recall that the directions $+\mathbf{e}_d$ and $-\mathbf{e}_d$ are physically equivalent in the films considered.

Let us express the electric field of the scattered field $\mathbf{E}_s$ in the far-field zone in the direction of $\mathbf{e}_s$ as [22]

$$\mathbf{E}_s = -\frac{\exp(ikr)}{-ikr}\frac{ik^3}{4\pi\varepsilon_p}\mathbf{e}_s \times [\mathbf{e}_s \times \mathbf{P}], \tag{2}$$

where *r* is the distance from the droplet center to the observation point.

In a rectangular system of coordinates $\mathbf{e}_\parallel\mathbf{e}_\perp\mathbf{e}_d$, where the unit vectors $\mathbf{e}_\perp$ and $\mathbf{e}_\parallel$ are directed along $\mathbf{e}_i \times \mathbf{e}_d$ and $\mathbf{e}_\perp \times \mathbf{e}_d$, the permittivity tensor $\underline{\varepsilon}$ has a diagonal form:

$$\underline{\varepsilon} = \begin{pmatrix} \varepsilon_{do} & 0 & 0 \\ 0 & \varepsilon_{do} & 0 \\ 0 & 0 & \varepsilon_{de} \end{pmatrix}, \tag{3}$$

where ε_{do} and ε_{de} are the effective permittivities of the LC droplet for the ordinary and extraordinary waves. To determine the dielectric permittivities of the LC droplet, we apply the effective-medium approximation [23]. Then,

$$\varepsilon_{do} = \varepsilon_{iso} - \frac{1}{3}\Delta\varepsilon S S_d, \tag{4}$$

$$\varepsilon_{de} = \varepsilon_{iso} + \frac{2}{3}\Delta\varepsilon S S_d. \tag{5}$$

Here,

$$\varepsilon_{iso} = \frac{2\varepsilon_o + \varepsilon_e}{3}, \quad \Delta\varepsilon = \varepsilon_e - \varepsilon_o,$$

ε_o and ε_e are the permittivities of the LC for ordinary and extraordinary waves and *S* is the molecular order parameter of the LC [3, 4, 21]. S_d is the order parameter of the LC droplet that characterizes the orientation ordering of the axes of LC molecules inside the droplet. In the isotropic phase with a chaotic orientation of molecules in the LC droplet, $S_d = 0$; when all the molecules are oriented along the same direction, $S_d = 1$.

Let us define the amplitude scattering matrix *S* with elements S_j, $j = 1, 2, 3, 4$, as

$$\begin{pmatrix} E_\parallel^s \\ E_\perp^s \end{pmatrix} = -\frac{\exp(ikr)}{ikr}\begin{pmatrix} S_2 & S_3 \\ S_4 & S_1 \end{pmatrix}\begin{pmatrix} E_\parallel^i \\ E_\perp^i \end{pmatrix}. \tag{6}$$

Here, $E_\parallel^i$ and $E_\perp^i$ are the components of the electric vector of the transmitted wave along the unit vectors $\mathbf{e}_\parallel^i$ and $\mathbf{e}_\perp^i$ ($\mathbf{e}_\perp^i \parallel [\mathbf{e}_i \times \mathbf{e}_s]$, $\mathbf{e}_\parallel^i \parallel [\mathbf{e}_\perp^i \times \mathbf{e}_i]$) and $E_\parallel^s$ and $E_\perp^s$ are the components of the electric vector of the scattered wave along the unit vectors $\mathbf{e}_\parallel^s$ and $\mathbf{e}_\perp^s$ ($\mathbf{e}_\perp^s \parallel [\mathbf{e}_i \times \mathbf{e}_s]$, $\mathbf{e}_\parallel^s \parallel [\mathbf{e}_\perp^s \times \mathbf{e}_s]$). Note that $\mathbf{e}_\perp^i = \mathbf{e}_\perp^s$.

Using expressions (1), (2), and (6) for the elements of the amplitude scattering matrix, we obtain

$$S_1 = -\frac{ik^3}{4\pi}\mathbf{e}_\perp^s\tilde{\underline{\underline{\alpha}}}\mathbf{e}_\perp^i, \quad (7)$$

$$S_2 = -\frac{ik^3}{4\pi}\mathbf{e}_\parallel^s\tilde{\underline{\underline{\alpha}}}\mathbf{e}_\parallel^i, \quad (8)$$

$$S_3 = -\frac{ik^3}{4\pi}\mathbf{e}_\parallel^s\tilde{\underline{\underline{\alpha}}}\mathbf{e}_\perp^i, \quad (9)$$

$$S_4 = -\frac{ik^3}{4\pi}\mathbf{e}_\perp^s\tilde{\underline{\underline{\alpha}}}\mathbf{e}_\parallel^i. \quad (10)$$

To determine the polarizability tensor $\tilde{\underline{\underline{\alpha}}}$ in the coordinate system $\mathbf{ee}'\mathbf{e}_i$, we write a transition matrix A from basis $\mathbf{ee}'\mathbf{e}_i$ to basis $\mathbf{e}_\parallel\mathbf{e}_\perp\mathbf{e}_d$, in which the polarizability tensor has a diagonal form:

$$A = \begin{pmatrix} \mathbf{e}_\parallel\cdot\mathbf{e} & \mathbf{e}_\parallel\cdot\mathbf{e}' & \mathbf{e}_\parallel\cdot\mathbf{e}_i \\ \mathbf{e}_\perp\cdot\mathbf{e} & \mathbf{e}_\perp\cdot\mathbf{e}' & \mathbf{e}_\perp\cdot\mathbf{e}_i \\ \mathbf{e}_d\cdot\mathbf{e} & \mathbf{e}_d\cdot\mathbf{e}' & \mathbf{e}_d\cdot\mathbf{e}_i \end{pmatrix}. \quad (11)$$

In the Rayleigh approximation, the relation between the polarizability tensor $\underline{\underline{\alpha}}$ and the permittivity tensor $\underline{\underline{\varepsilon}}$ for an optically soft LC droplet in the coordinate system $\mathbf{e}_\parallel\mathbf{e}_\perp\mathbf{e}_d$ can be expressed as

$$\underline{\underline{\alpha}} = v\left(\frac{\underline{\underline{\varepsilon}}}{\varepsilon_p} - \underline{\underline{I}}\right), \quad (12)$$

where v is the volume of the droplet and $\underline{\underline{I}}$ is a 3×3 identity matrix.

Let us divide the permittivity tensor $\underline{\underline{\varepsilon}}$ given by (3) into isotropic and anisotropic components [24]:

$$\underline{\underline{\varepsilon}} = \varepsilon_{iso}^d\underline{\underline{I}} - \frac{\Delta\varepsilon_d}{3}\underline{\underline{\beta}}, \quad (13)$$

where $\varepsilon_{iso}^d = (2\varepsilon_{do} + \varepsilon_{de})/3$, $\Delta\varepsilon_d = \varepsilon_{de} - \varepsilon_{do}$ is the optical anisotropy of the LC droplet and

$$\underline{\underline{\beta}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \quad (14)$$

Then, in the coordinate system $\mathbf{ee}'\mathbf{e}_i$, we have the following expression for the polarizability tensor $\tilde{\underline{\underline{\alpha}}}$ that appears in formulas (7)–(10):

$$\tilde{\underline{\underline{\alpha}}} = A^T\underline{\underline{\alpha}}A. \quad (15)$$

In view of the orthogonality of the bases $\mathbf{ee}'\mathbf{e}_i$ and $\mathbf{e}_\parallel\mathbf{e}_\perp\mathbf{e}_d$, we used here the transposed matrix A^T instead of the inverse matrix A^{-1}.

Let us express vectors $\mathbf{e}_\perp^{i,s}$, $\mathbf{e}_\parallel^s$, and $\mathbf{e}_\parallel^i$ in basis $\mathbf{ee}'\mathbf{e}_i$. Then, using formulas (7)–(15) for the elements of the amplitude matrix, after relevant mathematical transformations, we obtain

$$S_1 = -\frac{ik^3v}{4\pi}\left[\frac{\varepsilon_{do}}{\varepsilon_p} - 1 + \frac{\Delta\varepsilon_d}{\varepsilon_p}((\mathbf{e}_d\cdot\mathbf{e})(\mathbf{e}_\perp^s\cdot\mathbf{e}) + (\mathbf{e}_d\cdot\mathbf{e}')(\mathbf{e}_\perp^s\cdot\mathbf{e}'))^2\right], \quad (16)$$

$$S_2 = -\frac{ik^3v}{4\pi}\left[\left(\frac{\varepsilon_{do}}{\varepsilon_p} - 1\right)(\mathbf{e}_i\cdot\mathbf{e}_s) + \frac{\Delta\varepsilon_d}{\varepsilon_p}(\mathbf{e}_i\cdot\mathbf{e}_s)((\mathbf{e}_d\cdot\mathbf{e})(\mathbf{e}_\perp^s\cdot\mathbf{e}') - (\mathbf{e}_d\cdot\mathbf{e}')(\mathbf{e}_\perp^s\cdot\mathbf{e}))^2 + \frac{\Delta\varepsilon_d}{\varepsilon_p}(\mathbf{e}_d\cdot\mathbf{e}_i)(\mathbf{e}_\parallel^s\cdot\mathbf{e}_i)((\mathbf{e}_d\cdot\mathbf{e})(\mathbf{e}_\perp^s\cdot\mathbf{e}') - (\mathbf{e}_d\cdot\mathbf{e}')(\mathbf{e}_\perp^s\cdot\mathbf{e}))\right], \quad (17)$$

$$S_3 = -\frac{ik^3v}{4\pi}\frac{\Delta\varepsilon_d}{\varepsilon_p}[(\mathbf{e}_i\cdot\mathbf{e}_s) \times \{(\mathbf{e}_\perp^s\cdot\mathbf{e})(\mathbf{e}_\perp^s\cdot\mathbf{e}')((\mathbf{e}_d\cdot\mathbf{e})^2 - (\mathbf{e}_d\cdot\mathbf{e}')^2) + (\mathbf{e}_d\cdot\mathbf{e})(\mathbf{e}_d\cdot\mathbf{e}')((\mathbf{e}_\perp^s\cdot\mathbf{e}')^2 - (\mathbf{e}_\perp^s\cdot\mathbf{e})^2)\} + (\mathbf{e}_d\cdot\mathbf{e}_i)(\mathbf{e}_\parallel^3\cdot\mathbf{e}_i)((\mathbf{e}_d\cdot\mathbf{e})(\mathbf{e}_\perp^s\cdot\mathbf{e}) + (\mathbf{e}_d\cdot\mathbf{e}')(\mathbf{e}_\perp^s\cdot\mathbf{e}'))], \quad (18)$$

$$S_4 = -\frac{ik^3v}{4\pi}\frac{\Delta\varepsilon_d}{\varepsilon_p} \times [((\mathbf{e}_d\cdot\mathbf{e})^2 - (\mathbf{e}_d\cdot\mathbf{e}')^2)(\mathbf{e}_\perp^s\cdot\mathbf{e})(\mathbf{e}_\perp^s\cdot\mathbf{e}') + (\mathbf{e}_d\cdot\mathbf{e})(\mathbf{e}_d\cdot\mathbf{e}')((\mathbf{e}_\perp^s\cdot\mathbf{e}')^2 - (\mathbf{e}_\perp^s\cdot\mathbf{e})^2)]. \quad (19)$$

Equations (16)–(19) are expressed in the laboratory system of coordinates *xyz* without any assumption about the direction of illumination and the orientation of the droplet.

When the film is illuminated by a wave propagating parallel to the *z* axis ($\theta_i = 0$, $\mathbf{e}_i \parallel z$), vectors $\mathbf{e}$ and $\mathbf{e}'$ lie in the *xy* plane. In the laboratory system of coordinates *xyz*, for a fixed orientation of director $\mathbf{e}_d$ of a droplet, the

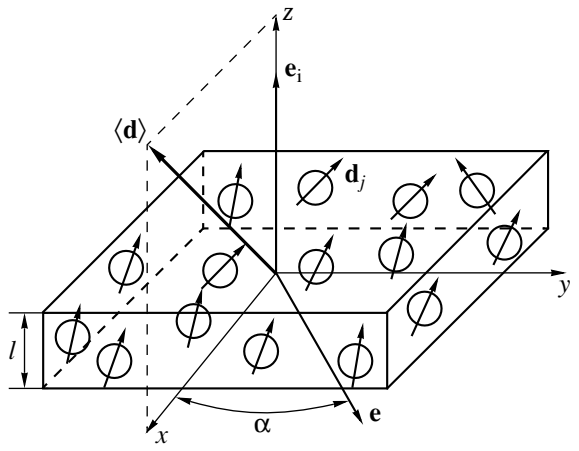


Fig. 2. Schematic view of a PDLC film and the geometry of its illumination; $\mathbf{e}_i$ is the direction of the incident wave, **e** is the direction of the polarization of the incident wave, α is the angle of polarization, $\mathbf{d}_j$ is the director of the jth droplet, $\langle \mathbf{d} \rangle$ is the orientation of the droplet directors, and l is the thickness of the film.

elements of the amplitude scattering matrix of the droplet are expressed as

$$S_1 = -\frac{ik^3 v}{4\pi}\left[\frac{\varepsilon_{do}}{\varepsilon_p} - 1 + \frac{\Delta\varepsilon_d}{\varepsilon_p}\sin^2\theta_d \sin^2(\varphi_d - \varphi_s)\right], \quad (20)$$

$$S_2 = -\frac{ik^3 v}{4\pi}\left[\left(\frac{\varepsilon_{do}}{\varepsilon_p} - 1\right)\cos\delta + \frac{\Delta\varepsilon_d}{\varepsilon_p}\cos\delta\sin^2\theta_d\cos^2(\varphi_d - \varphi_s) - \frac{\Delta\varepsilon_d}{2\varepsilon_p}\sin\delta\sin(2\theta_d)\cos(\varphi_d - \varphi_s)\right], \quad (21)$$

$$S_3 = -\frac{ik^3 v}{4\pi}\frac{\Delta\varepsilon_d}{2\varepsilon_p}[\cos\delta\sin^2\theta_d\sin(2(\varphi_d - \varphi_s)) - \sin\delta\sin(2\theta_d)\sin(\varphi_d - \varphi_s)], \quad (22)$$

$$S_4 = -\frac{ik^3 v}{4\pi}\frac{\Delta\varepsilon_d}{2\varepsilon_p}\sin^2\theta_d\sin(2(\varphi_d - \varphi_s)). \quad (23)$$

Here, δ is the scattering angle determined by vectors $\mathbf{e}_i$ and $\mathbf{e}_s$ (Fig. 1), φ_s is the angle between scattering plane $\mathbf{e}_i\mathbf{e}_s$ and the zx plane, and angle φ_d defines the orientation of the principal plane $\mathbf{e}_i\mathbf{e}_d$ with respect to the zx plane.

As a rule, PDLC films consist of polydispersed ensembles of LC droplets. Therefore, analysis of a coherent field transmitted through a PDLC film requires knowledge of the averaged elements of the scattering matrix. Let us average elements (20)–(23) of the scattering matrix over the sizes and orientations of droplets. Assuming that the distribution of the droplet directors is cylindrically symmetric, we obtain

$$\langle S_1\rangle = -\frac{ik^3\langle v\rangle}{4\pi}\left[\frac{\varepsilon_{do}}{\varepsilon_p} - 1 + \frac{\Delta\varepsilon_d}{3\varepsilon_p}(1 + 2S_x\sin^2\varphi_s + 2S_y\cos^2\varphi_s)\right], \quad (24)$$

$$\langle S_2\rangle = -\frac{ik^3\langle v\rangle}{4\pi}\left[\left(\frac{\varepsilon_{do}}{\varepsilon_p} - 1\right)\cos\delta + \frac{\Delta\varepsilon_d}{3\varepsilon_p}(1 + 2S_x\cos^2\varphi_s + 2S_y\sin^2\varphi_s)\cos\delta\right], \quad (25)$$

$$\langle S_3\rangle = \langle S_4\rangle = 0. \quad (26)$$

Here, the angle brackets denote averaging over the sizes of droplets and the orientations of their directors; $\langle v\rangle$ is the average volume of an LC droplet; and S_x, S_y, and S_z are order parameters that characterize the orientation ordering of the directors of LC droplets in the xyz coordinate system [16, 17]:

$$S_x = \frac{1}{2}(3\langle\sin^2\theta_d\cos^2\varphi_d\rangle - 1), \quad (27)$$

$$S_y = \frac{1}{2}(3\langle\sin^2\theta_d\sin^2\varphi_d\rangle - 1), \quad (28)$$

$$S_z = \frac{1}{2}(3\langle\cos^2\theta_d\rangle - 1). \quad (29)$$

The parameters S_x, S_y, and S_z are related by the equation

$$S_x + S_y + S_z = 0. \quad (30)$$

If $S_x = S_y$, then the elements (24) and (25) of the averaged amplitude scattering matrix coincide with the elements of the averaged amplitude scattering matrix obtained in [25].

The formulas for the elements of the amplitude scattering matrix that are derived in this section in the Rayleigh approximation can be used for analyzing the transmission of light through films containing particles such that the scattering by these particles is described in the Raylegh–Gans approximation. To this end, one should multiply each element (16)–(19) of the scattering matrix by an appropriate form factor [19, 22].

3. POLARIZATION TRANSFORMATION: BASIC RELATIONS

Consider a PDLC film under normal illumination by a linearly polarized plane wave. The laboratory system of xyz coordinates is shown in Fig. 2. The z axis defines the direction of the normal vector to the film, and the xy plane coincides with the lower boundary of the film. The x axis is chosen so that the director of the droplets, averaged over the volume of the film, $\langle \mathbf{d} \rangle$, lies in the zx plane. Then, a wave polarized along the x axis is

extraordinary, while a wave polarized along the y axis is ordinary.

Using the Foldy–Twersky equation, we determine the coherent field transmitted through the film [20, 22]. Let us write this equation for the vector case as

$$\begin{pmatrix} \langle E_e \rangle \\ \langle E_o \rangle \end{pmatrix} = \begin{pmatrix} \psi_e(z)|_{z=l} & 0 \\ 0 & \psi_o(z)|_{z=l} \end{pmatrix} \begin{pmatrix} \cos\alpha \\ \sin\alpha \end{pmatrix} E_i. \quad (31)$$

Here, $\langle E_o \rangle$ and $\langle E_e \rangle$ are the ordinary and extraordinary components of the coherent field of a wave transmitted through a PDLC film.

The functions $\psi_{e,o}(z)$ are solutions to the integral equations

$$\psi_{e,o}(z) = \exp(ikz) \times \left(1 - q\langle S_{e,o}(0)\rangle \int_0^z \exp(-ikz_s)\psi_{e,o}(z_s)dz_s\right), \quad (32)$$

where $q = 2\pi k^{-2}N_v$, N_v is the number of LC droplets per unit volume, and $\langle S_{e,o}(0)\rangle$ are the amplitude scattering functions, averaged over the droplet sizes and the orientations of their directors, for the extraordinary and ordinary waves for zero scattering angle ($\delta = 0$).

The solution to Eq. (32) is given by

$$\psi_{e,o}(z) = \exp(ik_{e,o}z). \quad (33)$$

Here, k_e and k_o are the propagation constants of the extraordinary and ordinary waves:

$$k_{e,o} = k + iq\langle S_{e,o}(0)\rangle. \quad (34)$$

The amplitude scattering functions $S_e(0)$ and $S_o(0)$ are calculated with the use of formula (25):

$$\langle S_e(0)\rangle = \langle S_2\rangle|_{\varphi_s=0,\,\delta=0} = -\frac{ik^3\langle v\rangle}{4\pi}\left[\frac{\varepsilon_{do}}{\varepsilon_p} - 1 + \frac{\Delta\varepsilon_d}{3\varepsilon_p}(1 + 2S_x)\right], \quad (35)$$

$$\langle S_o(0)\rangle = \langle S_1\rangle|_{\varphi_s=0,\,\delta=0} = -\frac{ik^3\langle v\rangle}{4\pi}\left[\frac{\varepsilon_{do}}{\varepsilon_p} - 1 + \frac{\Delta\varepsilon_d}{3\varepsilon_p}(1 + 2S_y)\right]. \quad (36)$$

Using formulas (31)–(36) for the components E_x and E_y defined as the real parts of the extraordinary and ordinary waves normalized by the field of the incident wave,

$$E_{x,y} = \mathrm{Re}\frac{\langle E_{e,o}\rangle}{E_i\exp(ikl)},$$

we obtain

$$E_x = t_e\cos\alpha\cos\Phi_e, \quad (37)$$

$$E_y = t_o\sin\alpha\cos\Phi_o, \quad (38)$$

$$\Phi_{e,o} = ql\,\mathrm{Im}\langle S_{e,o}(0)\rangle = -\pi\frac{l}{\lambda_p}c_v\left[\frac{\varepsilon_{do}}{\varepsilon_p} - 1 + \frac{\Delta\varepsilon_d}{3\varepsilon_p}(1 + 2S_{x,y})\right], \quad (39)$$

where $c_v = N_v\langle v\rangle$ is the volume concentration of LC droplets in the film. The coefficients t_e and t_o are defined by the relations

$$t_{e,o} = \exp(-\gamma_{e,o}l/2), \quad (40)$$

where γ_e and γ_o are the extinction indices of the extraordinary and ordinary waves, respectively:

$$\gamma_{e,o} = \sigma_{e,o}N_v, \quad (41)$$

$$\sigma_{e,o} = \frac{4\pi}{k^2}\mathrm{Re}\langle S_{e,o}(0)\rangle. \quad (42)$$

Note that, in the Rayleigh approximation, one cannot use formulas (42) to determine the total scattering cross sections σ_e and σ_o of nonabsorbing LC droplets because the real parts of the amplitude scattering functions $\langle S_e(0)\rangle$ and $\langle S_o(0)\rangle$ vanish [19, 22]. Therefore, one has to integrate the modules of the squares of the scattering matrix amplitudes over the total solid angle $\Omega = 4\pi$. Taking into account the fact that the amplitude scattering matrix is diagonal and that

$$\langle S_2\rangle|_{\varphi_s=0,\,\delta=0} = \langle S_1\rangle|_{\varphi_s=\pi/2},$$

$$\langle S_1\rangle|_{\varphi_s=0} = \langle S_2\rangle|_{\varphi_s=\pi/2,\,\delta=0},$$

we can write the following expressions for the total scattering cross sections:

$$\sigma_e = \frac{1}{k^2}\int_{4\pi}\left(\left|\langle S_2\rangle|_{\varphi_s=0}\right|^2 + \left|\langle S_1\rangle|_{\varphi_s=\pi/2}\right|^2\right)d\Omega, \quad (43)$$

$$\sigma_o = \frac{1}{k^2}\int_{4\pi}\left(\left|\langle S_1\rangle|_{\varphi_s=0}\right|^2 + \left|\langle S_2\rangle|_{\varphi_s=\pi/2}\right|^2\right)d\Omega. \quad (44)$$

Using (24), (25), and (41)–(44), we obtain the following formulas for the extinction indices γ_e and γ_o of spherical LC droplets:

$$\gamma_{e,o} = \frac{8}{9}\langle x\rangle^4 c_v f\langle d\rangle^{-1}\left(\frac{\varepsilon_{do}}{\varepsilon_p} - 1 + \frac{\Delta\varepsilon_d}{3\varepsilon_p}(1 + 2S_{x,y})\right)^2, \quad (45)$$

where $\langle x\rangle = \pi\langle d\rangle/\lambda_p$ is the mean diffraction parameter, $\langle d\rangle$ is the mean diameter of droplets, and f is the ratio of the third moment of the diameter distribution of droplets to the cube of the mean value of the diameter, f =

$\langle d^3\rangle/\langle d\rangle^3$. In the case of the gamma distribution of LC droplets with respect to their sizes [26, 27], $f = (1 + 2/\mu)(1 + 1/\mu)$, where μ is the distribution parameter.

To analyze the state of polarization of the transmitted wave, we determine the phase difference $\Delta\Phi$ between the ordinary and extraordinary waves. Using (4), (5), (13), and (39), we obtain

$$\Delta\Phi = \Phi_o - \Phi_e = \pi\frac{l}{\lambda_p}\frac{2\Delta\varepsilon}{3\varepsilon_p}SS_d(S_x - S_y). \quad (46)$$

Hence, a transformation of the state of polarization is only possible in the case of partial orientation of the droplet directors ($S_x \neq S_y$). For a full orientation of the droplet directors (along the x axis), $S_x = 1$, $S_y = S_z = -1/2$, and the difference $S_x - S_y = 3/2$ becomes maximal. If $S_x = S_y$ (cylindrically symmetric distribution of droplet directors with respect to the z axis or a chaotic distribution of directors), then $\Delta\Phi = 0$; i.e., the original state of polarization is preserved. In this case, the problem of wave propagation through a PDLC film under normal illumination can be considered in the scalar approximation [25].

In the system of coordinates chosen, we have $S_x > S_y$ (Fig. 2). The order parameters S_x and S_y can be determined from (27)–(30):

$$S_x = \frac{1}{2}((1 - S_z)g - S_z), \quad (47)$$

$$S_y = \frac{1}{2}((S_z - 1)g - S_z), \quad (48)$$

$$S_x - S_y = g(1 - S_z), \quad (49)$$

$$g = \langle\cos^2\varphi_d\rangle - \langle\sin^2\varphi_d\rangle. \quad (50)$$

Suppose that we have a uniform angular distribution of the probability density φ_d. Then,

$$g = \mathrm{sinc}(2\varphi_{dm}), \quad (51)$$

where φ_{dm} is the angle of maximal deviation of the droplet director from the x axis.

Formulas (47)–(51) allow one to analyze the state of polarization as a function of the order parameters of the PDLC film during the transition from a partially ordered structure of droplet directors to a homeotropic structure, where the directors of all droplets (with positive anisotropy of the LC, $\Delta\varepsilon > 0$) are directed along the control field (along axis z). The direction of the averaged director of droplets $\langle d\rangle$ depends on the magnitude of the applied field. As the magnitude of the applied field increases, the angle between the z axis and vector $\langle\mathbf{d}\rangle$ decreases. This angle vanishes in the limit of infinite field. The transition from a partially ordered structure of droplet directors to a homeotropic structure differs from the transition from a chaotic to a homeotropic structure, which occurs in most PDLC films, by the range of variation of order parameter S_z. In the former case, $-1/2 \le S_z \le 1$, whereas, in the latter, $0 \le S_z \le 1$.

To analyze the polarization characteristics of the transmitted light, let us write an equation for the polarization ellipse using formulas (37) and (38):

$$\frac{E_x^2}{a_e^2} - \frac{2E_xE_y}{a_ea_o}\cos\Delta\Phi + \frac{E_y^2}{a_o^2} = \sin^2\Delta\Phi, \quad (52)$$

where $a_e = t_e\cos\alpha$ and $a_o = t_o\sin\alpha$. The semiaxes A and B of the polarization ellipse and the rotation angle ξ of the axes of the polarization ellipse with respect to the xy axes of the laboratory system of coordinates are given by

$$A^2 = a_e^2\cos^2\xi + a_o^2\sin^2\xi + a_ea_o\sin(2\xi)\cos\Delta\Phi, \quad (53)$$

$$B^2 = a_e^2\sin^2\xi + a_o^2\cos^2\xi - a_ea_o\sin(2\xi)\cos\Delta\Phi, \quad (54)$$

$$\tan(2\xi) = 2a_ea_o\frac{\cos\Delta\Phi}{a_e^2 - a_o^2}. \quad (55)$$

Formulas (53)–(55) allow one to analyze the ellipsometric parameters (azimuth and ellipticity) of the light transmitted through a PDLC film during the transition from a partially oriented structure of droplet directors to a homeotropic structure and to determine the sign of rotation of the polarization plane and the conditions under which the transmitted wave is circularly polarized.

4. ANALYSIS OF THE RESULTS

Let us define the azimuth ξ_{ell} of the polarization ellipse as the angle between the major axis of the ellipse and the positive direction of the x axis, measured counterclockwise with respect to the positive direction of the z axis. Define the ellipticity η as the ratio of the minor and major axes of the polarization ellipse.

Consider films in which the phase difference $\Delta\Phi$ of the transmitted wave changes from π to zero under the transition from a partially ordered structure of droplet directors to a homeotropic structure.

It follows from (53)–(55) that, for $\Delta\Phi = \pi$, a wave transmitted through a film at an angle of β ($\tan\beta = (t_o/t_e)\tan\alpha$) is linearly polarized (line *1* in Fig. 3). For $\Delta\Phi = 0$, the transmitted wave retains the linear polarization of the incident wave (line *5* in Fig. 3). When $\pi > \Delta\Phi > 0$, the transmitted wave is elliptically polarized. The polarization ellipse rotates clockwise for an optical anisotropy of $\Delta\varepsilon > 0$ (ellipses *2*, *3*, and *4* in Fig. 3). The maximal ellipticity is obtained for $\Delta\Phi = \pi/2$ (ellipse *3* in Fig. 3). When $\Delta\Phi = \pi/2$ and a wave is incident at an angle of α_o ($\tan\alpha_o = t_e/t_o$), then the transmitted wave is circularly polarized ($\eta = 1$). Figure 3 illustrates the rotation of the polarization ellipse and of the plane of

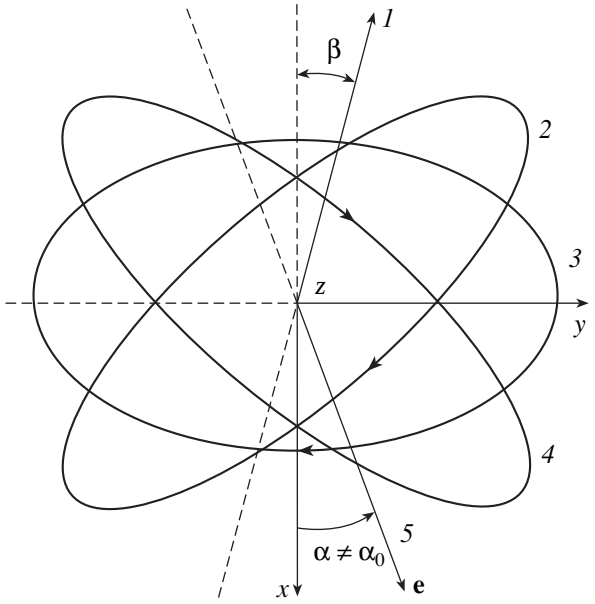


Fig. 3. Schematic view of the shape and orientation of the polarization ellipse. The arrows indicate the sign of rotation of the electric vector. Notations are explained in the text.

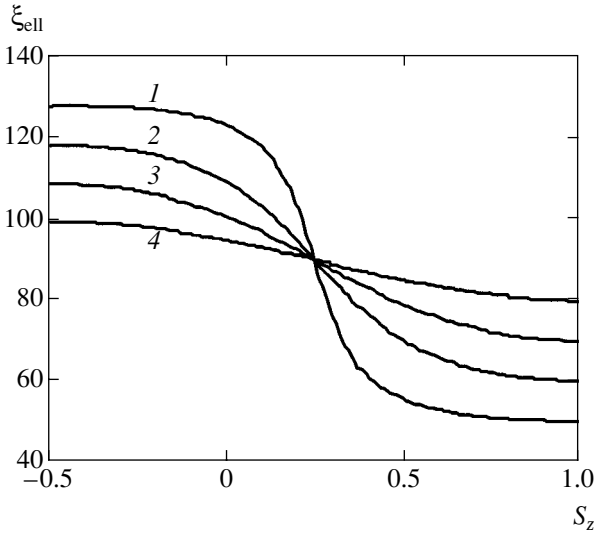


Fig. 4. Azimuth of the polarization ellipse of the transmitted light, ξ_{ell}, as a function of order parameter S_z for different polarization angles of the incident linearly polarized light: (*1*) $\alpha = 50°$, (*2*) 60°, (*3*) 70°, and (*4*) 80°. Parameters of the film are given in the text.

polarization of the transmitted wave for polarization angles $\alpha \neq \alpha_o$ of the incident wave. The electric vector rotates clockwise from the positive direction of the z axis (right elliptic polarization [19]).

It follows from (46) and (49) that the phase difference $\Delta\Phi$ is a function of the order parameter S_z. The azimuth ξ_{ell} of the polarization ellipse and the ellipticity η are shown in Figs. 4 and 5 versus the order parameter S_z (which determines the phase difference $\Delta\Phi$) for various polarization angles α of the incident wave. The calculations are performed by formulas (46) and (53)–(55) for refractive indices of an LC of $n_o = 1.511$ and $n_e = 1.74$ ($n_o^e = \varepsilon_o$ and $n_e^2 = \varepsilon_e$); a refractive index of the polymer of $n_p = 1.524$; order parameters of $S = 0.6$ and $S_d = 0.7$; a maximal deviation angle of the droplet director from the x axis of $\varphi_{dm} = 5°$; a mean diameter of LC droplets of $\langle d \rangle = 75$ nm; a parameter of gamma distribution of $\mu = 15$; a volume concentration of LC droplets of $c_v = 0.075$; a film thickness of $l = 41.3$ μm; a wavelength of the incident wave of $\lambda = 0.6328$ μm; and polarization angles of the incident wave of $\alpha = 50°$, 60°, 70°, and 80°.

The maximal rotation angle of the plane of polarization is defined as the difference of azimuths of the polarization ellipse for the order parameters $S_z = -1/2$ and $S_z = 1$. In the case under consideration, these order parameters correspond to a phase difference $\Delta\Phi$ equal to π and 0, respectively, for which the transmitted wave retains its linear polarization (Fig. 5). A change in the phase difference from π to zero can also be achieved for a smaller range of variation of the order parameter S_z. This allows one to solve the optimization problem on the rotation of the polarization plane under a change of the structure of the droplet directors in a PDLC film.

Derivation of Expression for the Order Parameter S_z

In practical applications, one has to know the order parameter S_z as a function of the control field. Such functions are given in [4, 21] for films in which a cha-

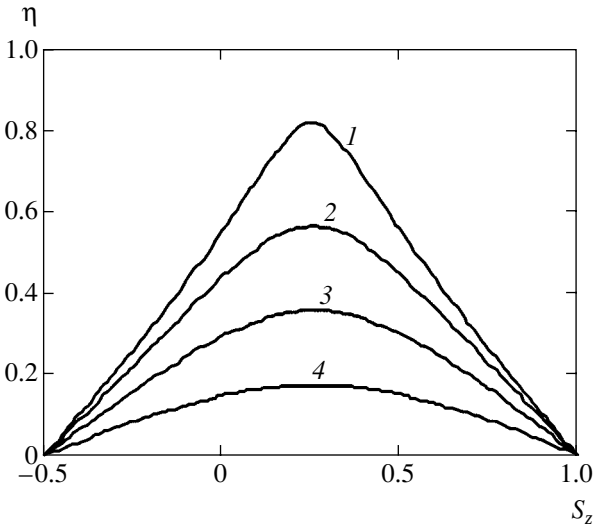


Fig. 5. Ellipticity η as a function of order parameter S_z for various polarization angles: (*1*) $\alpha = 50°$, (*2*) 60°, (*3*) 70°, and (*4*) 80°. Parameters of the film are given in the text.

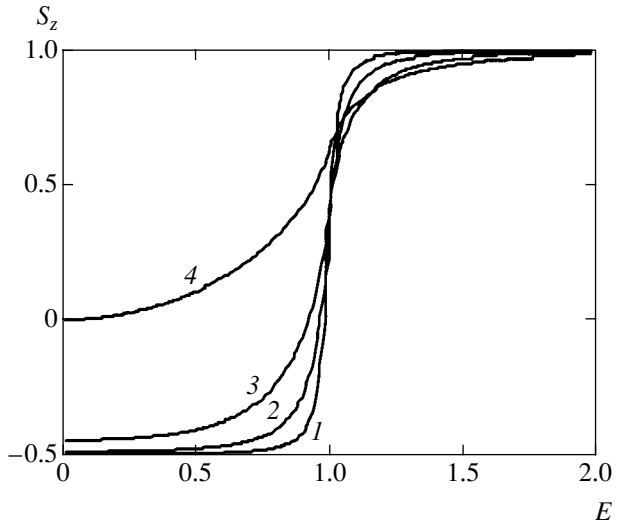


Fig. 6. Order parameter S_z as a function of normalized control field E. Transition from a partially ordered structure of droplet directors in a PDLC film to a homeotropic structure for (*1*) $\theta_m = 5°$, (*2*) 10°, and (*3*) 20°. (*4*) Transition from a chaotic structure of directors to a homeotropic structure for $\theta_m = 90°$.

otic structure of droplet directors is transformed into a homeotropic structure.

We investigate a transition of a partially oriented structure of LC droplet directors in a PDLC film to a homeotropic structure. Let us derive an expression for the order parameter in such a system. Let us apply the relations

$$S_z = \frac{1}{4} + \frac{3}{4}\langle \cos(2\theta)\rangle, \tag{56}$$

$$\cos(2\theta) = \frac{E^2 - 1 + 2\cos^2\theta_0}{\sqrt{(E^2-1)^2 + 4E^2\cos^2\theta_0}}. \tag{57}$$

Here, E is the control electric field normalized by a threshold value, θ is the angle between the director of a droplet and the z axis along which the control field is applied, θ_0 is the angle between the director of a droplet and the z axis for $E = 0$, and angular brackets denote averaging over the angle θ_0.

Suppose that the angular distribution of the droplet directors is uniform within a solid angle $\Delta\Omega$, $-\varphi_{dm} \le \varphi_d \le \varphi_{dm}$, $\pi/2 - \theta_m \le \theta_0 \le \pi/2 + \theta_m$, where φ_{dm} and θ_m are the maximal deviation angles of the droplet directors from the x axis in the xy and xz planes, respectively. Then,

$$\langle \cos(2\theta)\rangle = \frac{1}{4\varphi_{dm}\sin\theta_m} \times \int_{-\varphi_{dm}}^{\varphi_{dm}} d\varphi \int_{\pi/2-\theta_m}^{\pi/2+\theta_m} \cos(2\theta)\sin\theta_0 d\theta_0. \tag{58}$$

After integrating, we obtain

$$\langle \cos(2\theta)\rangle = a_1 + \frac{a_2 \ln a_3}{\sin\theta_m}, \tag{59}$$

$$a_1 = \frac{u}{4E^2}, \tag{60}$$

$$a_2 = \frac{3E^4 - 2E^2 - 1}{16E^3}, \tag{61}$$

$$a_3 = \left|\frac{2E\sin\theta_m + u}{u - 2E\sin\theta_m}\right|, \tag{62}$$

$$u = \left((E^2-1)^2 + 4E^2\sin^2\theta_m\right)^{1/2}. \tag{63}$$

If $\theta_m = \pi/2$, then formulas (56) and (59)–(63) for order parameter S_z are reduced to the well-known relations [4, 21] that describe the transition from a chaotic structure of droplet directors to a homeotropic structure.

Order parameter S_z as a function of control field E is shown in Fig. 6 for different angles θ_m. One can see that the transition from a partially ordered structure of droplet directors to a homeotropic structure is characterized by stronger dependence of the order parameter on the control field and offers more possibilities for modulating optical radiation than the transition from a chaotic to a homeotropic structure.

5. CONCLUSIONS

The method proposed in this paper enables one to analyze the polarization state of a wave transmitted through composite LC systems in the case of normal incidence of a linearly polarized wave. Examples of these systems may be given by dispersed LC films with nanosized nematic droplets, polymer networks, porous glasses, and other structures with nanosized LC elements [5, 28–32]. This method is a development of the method, accounted by the present authors in [25], for describing a polarization-independent phase modulation of light [12] transmitted through PDLC films.

The results obtained relate the morphological characteristics of a film to its electro-optic response. They are of interest in designing new types of electrically and magnetically controlled polarizers based on composite LC materials.

REFERENCES

1. L. Lucchetti and F. Simoni, J. Appl. Phys. **88**, 3934 (2000).
2. K. Amundson, A. van Blaaderen, and P. Wiltzius, Phys. Rev. E **55**, 1646 (1997).
3. G. M. Zharkova and A. S. Sonin, *Mesomorphic Composites* (Nauka, Moscow, 1994) [in Russian].

4. F. Simoni, *Nonlinear Optical Properties of Liquid Crystals and Polymer Dispersed Liquid Crystals* (World Sci., Singapore, 1997).
5. *Liquid Crystals in Complex Geometries*, Ed. by G. P. Crawford and S. Zumer (Taylor and Francis, London, 1996).
6. V. Ya. Zyryanov, E. P. Pozhidaev, S. L. Smorgon, *et al.*, Liq. Cryst. **28**, 433 (2001).
7. R. L. Sutherland, V. P. Tondiglia, L. V. Natarajan, *et al.*, Appl. Phys. Lett. **64**, 1074 (1994).
8. C. C. Bowley, P. A. Kossyrev, G. P. Crawford, and S. Faris, Appl. Phys. Lett. **79**, 9 (2001).
9. D. S. Wiersma, S. Gottardo, R. Sapienza, *et al.*, in *Wave Scattering in Complex Media: from Theory to Applications*, Ed. by B. A. van Tiggelen and S. E. Skipetrov (Kluwer Academic, Dordrecht, 2003), pp. 3–20.
10. M. J. Sansone, G. Khanarian, and T. M. Leslie, J. Appl. Phys. **67**, 4253 (1990).
11. M. Jazbinsek, I. Drevensek-Olenik, M. Zgonik, *et al.*, J. Appl. Phys. **90**, 3831 (2001).
12. L. Lucchetta, R. Karopiran, A. Mann, and F. Simoni, J. Appl. Phys. **91**, 6060 (2002).
13. S. Matsumoto, K. Hirabayashi, S. Sakata, *et al.*, IEEE Photon. Tech. Lett. **11**, 442 (1999).
14. S. Matsumoto, Y. Sugiyama, S. Sakata, *et al.*, Liq. Cryst. **27**, 649 (2000).
15. W. A. Crossland, T. D. Wilkinson, I. G. Manolis, *et al.*, Mol. Cryst. Liq. Cryst. **375**, 1 (2002).
16. L. M. Blinov, *Electro-Optical and Magneto-Optical Properties of Liquid Crystals* (Nauka, Moscow, 1978; Wiley, New York, 1983).
17. I.-C. Khoo, *Liquid Crystals. Physical Properties and Nonlinear Optical Phenomena* (Wiley, New York, 1995).
18. S. Chandrasekhar, *Liquid Crystals*, 2nd ed. (Cambridge Univ. Press, Cambridge, 1992; Mir, Moscow, 1980).
19. C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles* (Wiley, New York, 1983; Mir, Moscow, 1986).
20. A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978; Mir, Moscow, 1994), Vol. 2.
21. J. R. Kelly and P. Palffy-Muhoray, Mol. Cryst. Liq. Cryst. **243**, 11 (1994).
22. H. C. van de Hulst, *Light Scattering by Small Particles* (Wiley, New York, 1957; Inostrannaya Literatura, Moscow, 1961).
23. V. A. Loiko, A. V. Konkolovich, F. Simoni, *et al.*, in *Proceedings of 10th SID Symposium on Advanced Display Technologies* (Minsk, 2001), p. 58.
24. S. Zumer and J. W. Doane, Phys. Rev. A **34**, 3373 (1986).
25. V. A. Loĭko and A. V. Konkolovich, Zh. Éksp. Teor. Fiz. **123**, 552 (2003) [JETP **96**, 489 (2003)].
26. O. A. Afonin, Yu. V. Panina, A. B. Pravdin, and D. A. Yakovlev, Liq. Cryst. **15**, 395 (1997).
27. V. A. Loiko and A. V. Konkolovich, J. Phys. D: Appl. Phys. **33**, 2201 (2000).
28. T. L. Bunning, L. V. Natarajan, V. P. Tondiglia, *et al.*, Polymer **37**, 3147 (1996).
29. D. R. Cairns, C. C. Bowley, S. Danworaphong, *et al.*, Appl. Phys. Lett. **77**, 2677 (2000).
30. S. D. Hudson, H.-T. Jung, P. Kewsuwan, *et al.*, Liq. Cryst. **26**, 1493 (1999).
31. I. Dierking, M. A. Osipov, and S. T. Lagerwall, Eur. Phys. J. E **2**, 303 (2000).
32. A. Glushchenko, H. Kresse, V. Reshetnyak, *et al.*, Liq. Cryst. **23**, 241 (1997).

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 99, No. 2, 2004, pp. 352–362.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 2, 2004, pp. 395–406.

SOLIDS
Electronic Properties

The Influence of Kinetic Processes at a Negative Muon Track Endpoint in Doped Nondegenerate Silicon on the Behavior of Muon Spin Polarization

A. S. Baturin, V. N. Gorelkin, and V. R. Solov'ev

Moscow Institute of Physics and Technology, Institutskiĭ per. 9, Dolgoprudnyĭ, Moscow oblast, 141700 Russia
e-mail: baturin@lafeet.mipt.ru

Received November 20, 2003

Abstract—This work is aimed at creating a theoretical basis for analyzing and interpreting the results of μSR experiments in doped nondegenerate semiconductors at temperatures below 50 K, when the influence of kinetic processes at the endpoint of a muon stopping track on the behavior of its polarization is substantial. The effects related to the formation of free electrons and holes in a solid-state plasma at a muon track endpoint are shown to be responsible for the mere possibility of observing negative muon spin precession at the muon frequency in doped nondegenerate semiconductors at low temperatures. The Vangsness–Bloch equations are generalized to the case of parameters varying with time. A theory based on these generalized equations allowed us to interpret more correctly the available experimental results of μSR studies of semiconductors with the use of negative muons. We showed that the μSR method could be used to obtain information about the cross sections of exchange scattering of electrons and holes by impurity centers in the region of energies inaccessible to the other measurement techniques and to estimate the cross section of capture by a solitary charged Coulomb center at virtually all charge carrier concentrations and temperatures. Under the conditions when the Debye radius is larger than the mean distance between charged particles but smaller than the Thomson radius, the capture (recombination) cross section is described by a temperature dependence qualitatively different from that predicted by current theory. *© 2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

A negative muon stopped in silicon is captured by an orbit close to the silicon nucleus, decreases the effective charge of the nucleus by one, and thereby forms the so-called muonic aluminum atom $_{\mu}$Al, which remains in the lattice site and is an acceptor center in a silicon semiconductor with diamond structure [1]. During muon deceleration and capture, ions, electrons, and holes are formed that produce a solid-state plasma region at the muon track endpoint.

As in a pure semiconductor, there are acceptor center states of two types in *n*-type silicon, neutral paramagnetic $_{\mu}Al^0$ (*p*) and negatively charged diamagnetic $_{\mu}Al^-$ (*d*). The diamagnetic state is more favorable energetically in *n*-type silicon and is largely formed from the paramagnetic state as a result of electron capture. In μSR studies, these two states or kinds of systems are usually called fractions.

The electronic moment of the shell of an acceptor center is zero in the diamagnetic state, and this is why we observe nearly undamped precession at the free muon frequency $\omega_d = \omega_\mu$.

The electronic moment *j* of the shell of an acceptor center is nonzero in the paramagnetic state ($j = 3/2$ for $_{\mu}Al^0$). Under the conditions of experiments performed in [2–11], the electronic moment relaxation rate ν_e of the acceptor center was much higher than the effective hyperfine structure constant *A*. The hyperfine interaction of the electronic moment with the muon spin (acceptor center nucleus spin) therefore shifted the frequency of muon spin precession $\omega_p = \omega_\mu + \delta\omega$ in an external magnetic field with respect to the free muon precession frequency ω_μ.

Precession damping in the paramagnetic state is caused by two processes. First, the electronic moment of the acceptor center shell can relax in interactions with lattice phonons as a result of the exchange interaction between the acceptor center and neighboring nonionized impurities and exchange scattering of free carriers. The second reason for a decrease in the precession amplitude of muon polarization in the paramagnetic state is the transition of the acceptor center from the paramagnetic into diamagnetic state at a rate of ν_{tr}.

The frequency shift of muon precession was measured by the μSR method in silicon samples with various concentrations of dopants [2–11]. Simultaneously, the rate of precession damping and the precession amplitude were determined. All this information is con-

tained in the time histogram of the number of muon decays accumulated during measurements; this histogram is called the μSR signal. It should be borne in mind in analyzing experimental data that the characteristic muon spin precession frequency in magnetic fields on the order of 1kG is about 10^8 s^{-1}; for this reason, the "initial" time of μSR measurements is approximately 10^{-8} s after the formation of $_\mu$Al.

Experimental data are usually processed using the Vangsness–Bloch equations, which describe changes in muon spin polarization, on the assumption that the rate ν_e of electronic moment relaxation and the transition rate ν_{tr} are constant in time [7]. In the solid-state plasma formed in the region of muon localization, the ν_e and ν_{tr} values depend on the nonstationary concentration of electrons and holes. For this reason, the kinetic processes at the muon track endpoint must be taken into account to correctly interpret the totality of experimental results. This entails generalizing the Vangsness–Bloch equations to the case when the rates of relaxation and transition depend on time. This is the problem addressed in the present work; we consider the conditions of μ^-SR experiments with nondegenerate doped silicon samples.

2. A GENERALIZATION OF THE VANGSNESS–BLOCH EQUATIONS

Let us make a terminological comment at the outset. By the Vangsness–Bloch equations, we understand the linear equations that describe the relaxation of a quantum-mechanical system derived on the assumption that back action of the system on the thermostat (lattice) is absent. This derivation is based on the method of random trajectories and goes beyond the scope of perturbation theory. Generally, the behavior of the polarizations $P_i(a, t)$ that describe a system of kind a is determined by the Korst equations [12], with which one can acquaint oneself the most simply in monograph [13]. These equations read

$$\frac{d}{dt}P_i(a,t) = \gamma_{ik}(a,t)P_k(a,t) - \nu(a,t)P_i(a,t) + \sum_{a'\neq a}\nu(a'\longrightarrow a,t)P_i(a',t), \quad (1)$$

where

$$\nu(a,t) = \sum_{a'\neq a}\nu(a\longrightarrow a',t)$$

is the summed frequency of transitions from fraction a into all other fractions, $\nu(a'\longrightarrow a, t)$ is the frequency of transitions from fraction a' into fraction a, and $\gamma_{ik}(a, t)$ is the tensor that determines the time behavior of polarization in fraction a. In the particular case of precession at the frequency ω_p with respect to the magnetic field direction given by the vector $\mathbf{q}$ and damping, whose rate is different for different polarization components and is determined by the tensor Λ_{ik}, the $\gamma_{ik}(a, t)$ tensor has the form

$$\gamma_{ik} = \omega_p e_{ilk}q_l + \Lambda_{ik} - \Lambda_{ll}\delta_{ik}, \quad (2)$$

where δ_{ik} is the Kronecker symbol and e_{ilk} is the third-rank antisymmetric unit tensor.

Generally, index i may denote the number of the orthogonal operator whose coefficient is $P_i(a, t)$ in the expansion of the density matrix of system a in such operators.

If there are no transitions between the fractions and $\nu(a'\longrightarrow a, t) = 0$, it follows from (1) that the evolution of the polarization $P_i(a, t)$ in a fraction of kind a is determined by the linear equation

$$\frac{dP_i(a,t)}{dt} = \gamma_{ik}(a,t)P_k(a,t), \quad (3)$$

whose general solution can be represented as

$$P_i(a,t) = \mu_{ik}(a,t,t')P_k(a,t'), \quad (4)$$

where the evolution tensor $\mu_{ik}(a, t, t')$ is unambiguously determined by the $\gamma_{ik}(a, t)$ tensor,

$$\frac{d\mu_{il}(a,t,t')}{dt} = \gamma_{ik}(a,t)\mu_{kl}(a,t,t'). \quad (5)$$

For time-dependent transition rates $\nu(a'\longrightarrow a, t)$ from fraction a' to a, (1) can be written in the integral form

$$P_i(a,t) = \exp\left\{-\int_0^t \nu(a,\tau)d\tau\right\}\mu_{ik}(a,t,0)P_k(a,0) + \sum_{a'\neq a}\int_0^t \exp\left\{-\int_{t'}^t \nu(a,\tau)d\tau\right\}\mu_{ik}(a,t,t') \times \nu(a'\longrightarrow a,t')P_k(a',t')dt'. \quad (6)$$

The validity of this representation can be proved by differentiating (6) with respect to time taking into account (4) and (5).

Equations (5) and (6) are generalizations of the well-known equations for positive muon spin depolarization in the muonium atom (electronic moment 1/2) with a constant "electronic spin flip frequency" ν_e. For the first time, these equations were solved for several limiting cases in [14]. A fairly complete theory of the mechanism of muonium depolarization in matter was developed in [15–18]. A detailed treatment of this problem and bibliography can be found in monograph [19].

Equations (1)–(5) describe muonium-like systems with arbitrary electron shell moments. Relaxation

equations for such systems in the absence of transitions between fractions were considered in [20]. Two kinds of fractions exist for the $_{\mu}$Al acceptor center: diamagnetic, in which the electron shell surrounding the muon is completely filled, and paramagnetic with the total effective moment $j = 3/2$ and the hyperfine interaction constant A between the muon spin and electronic moment. A description of the paramagnetic fraction at electronic moment relaxation rates lower than or comparable to the hyperfine splitting constant must be based on relaxation equations similar to those considered in [20]. Such a situation arises in silicon at temperatures of several K, when damping on the order of the precession period is observed experimentally at the free muon frequency. Currently, this situation is quite accessible to measurements and theoretical analysis. However, traditionally, higher temperatures are studied, when the electronic moment relaxation rate is much higher than the hyperfine splitting constant. Hyperfine fields are then averaged, and muon spin precession occurs at a frequency shifted with respect to the precession frequency in the diamagnetic fraction because of the magnetization of the electron shell by an external field. Both diamagnetic and paramagnetic fractions can then be characterized by a precessing vector of muon spin polarization. In what follows, we consider precisely this situation. For this reason, all values with indices (except index *a*, which specifies the kind of the system) in the equations given above can be treated as usual Cartesian vectors and tensors.

If the relaxation mechanism is isotropic, the relaxation tensor $\Lambda_{ik}(p, t)$ in Eq. (2) for the paramagnetic fraction is given by two scalar functions (the relaxation rates for the muon spin polarization components longitudinal $\Lambda_1(t)$ and transverse $\Lambda_2(t)$ to the magnetic field) and the unit vector q_i aligned with the external magnetic field,

$$\Lambda_{ik}(p, t) = q_i q_k \Lambda_1(t) + (\delta_{ik} - q_i q_k)\Lambda_2(t). \quad (7)$$

Substituting (7) into (2) and solving (5) with the resulting γ_{ik} tensor yields the μ_{ik} tensor in the form

$$\mu_{ik}(p, t, t') = q_i q_k \exp\left(-\int_{t'}^{t} \Lambda_1(\tau)d\tau\right)$$
$$+ \{(\delta_{ik} - q_i q_k)\cos\omega(t - t') + e_{ikl}q_l \sin\omega(t - t')\} \quad (8)$$
$$\times \exp\left(-\int_{t'}^{t} \Lambda_2(\tau)d\tau\right).$$

In the absence of transitions between the fractions and when the relaxation rate of the electronic moment of the acceptor center shell in the paramagnetic state is independent of time and much larger than the effective hyperfine interaction constant A ($\nu_e \gg A$), the shift $\delta\omega$ of the precession frequency and the Λ_2 rate of precession damping for the paramagnetic fraction are given by the equations [21]

$$\frac{d\omega}{\omega_d} = -A\frac{5g_1\mu_B\hbar}{4\mu_{\mu B}T},$$
$$\Lambda_2 = \frac{5A^2}{4\nu_e}\left(1 + \frac{1}{1 + (\omega_e/\nu_e)^2}\right). \quad (9)$$

Here, μ_B and $\mu_{\mu B}$ are the Bohr magnetons for the electron and muon, respectively; T is the temperature; and $\omega_e = g_1\mu_B B/\hbar$ is the electronic moment precession frequency in magnetic field B. In this work, we used $g_1 = -0.98$ for $_{\mu}$Al [22]. The effective hyperfine interaction constant A is determined by a linear combination of the spin Hamiltonian constants [23]. The currently available data of μ^-SR experiments on the paramagnetic shift in silicon give an estimate of $A \approx 25$ MHz but cannot be used to obtain unambiguous estimates for the anisotropic and isotropic parts of the hyperfine interaction constant thus far. Obtaining such data can be made possible by studying the strain dependence of the paramagnetic shift.

In the absence of transitions between the fractions and if the rate ν_e of electronic moment relaxation is independent of time, solution (4) predicts exponential damping of paramagnetic fraction polarization $P_i(p, t)$ at a constant rate Λ_2. Dependence of the concentration of charge carriers on time can, however, result in a nonexponential dependence of muon spin polarization $P_i(p, t)$ relaxation, because the tensor of relaxation $\Lambda_{ik}(p, t)$ then becomes a function of time.

After electron capture by a neutral acceptor center in the paramagnetic state and the formation of diamagnetic $_{\mu}Al^-$, the reverse process accompanied by hole capture does not occur ($\nu_{d \to p} = 0$), because the forbidden bandwidth (≈2 eV) in the temperature range under consideration (below several dozen K) is much larger than T and the concentration of holes is close to zero. In the absence of reverse transitions, the relative population of the paramagnetic fraction $w(p, t)$ satisfies the equation

$$\frac{dw(p, t)}{dt} = -\nu_{tr}(t)w(p, t), \quad (10)$$

whose solution is

$$w(p, t) = \exp\left\{-\int_{0}^{t} \nu_{tr}(\tau)d\tau\right\}. \quad (11)$$

As $\nu_{d \to p} = 0$, it follows from general equation (6) for the paramagnetic fraction that

$$P_i(p, t) = w(p, t)\mu_{ik}(p, t)P_k(p, 0). \quad (12)$$

We assume for definiteness that the direction of the z axis coincides with the external magnetic field direc-

tion and the initial polarization is perpendicular to z and coincides with the direction of the x axis. Following tradition, the projection of polarization onto the x axis will be called transverse component $P_\perp$. Using (8) and (11) in (12) yields the transverse polarization component in the form

$$P_\perp(p, t) = P_\perp(p, 0)w(p, t)F(p, t)\cos(\omega_p t), \quad (13)$$

where

$$F(p, t) = \exp\left\{-\int_0^t \Lambda_2(\tau)d\tau\right\}.$$

For the diamagnetic fraction, the general equation (6) gives

$$P_\perp(d, t) = P_\perp(d, 0)\cos(\omega_d t) + \int_0^t \cos(\omega_d(t-t'))\nu_{tr}(t')P_\perp(p, t')dt'. \quad (14)$$

The $p \longrightarrow d$ transition contributes to both relaxing $P_\perp(p, t)$ and nonrelaxing $P_\perp(d, t)$ polarization components. Their precession amplitudes cannot therefore be identified with the corresponding fraction populations.

The initial polarizations in (13) and (14) are determined by the concentration of donor impurities N_d. The diamagnetic state $_\mu Al^-$ with a captured electron is equilibrium in n-type silicon. If a donor is situated in the immediate vicinity of the $_\mu Al$ acceptor center, than the transition of the donor–acceptor pair to the state in which the donor loses its electron and the acceptor is formed in the diamagnetic state $_\mu Al^-$ can be favored. The distance R between donor and acceptor at which a donor–acceptor pair is formed is found from the condition that the interaction energy between an electron and the acceptor center be comparable with the ionization energy of the donor [24]. This distance estimated for silicon is $R \approx 10^{-8}/\varepsilon_d$, where ε_d is the donor ionization energy in eV. The probability for a donor to occur inside a sphere of radius R and thereby form $_\mu Al^-$ in the diamagnetic state, which corresponds to the initial polarization $P_\perp(d, 0)$, is

$$P_\perp(d, 0) = 1 - \exp\left(-\frac{4}{3}\pi R^3 N_d\right). \quad (15)$$

Accordingly,

$$P_\perp(p, 0) = 1 - P_\perp(d, 0). \quad (16)$$

3. THE BEHAVIOR OF TRANSVERSE MUON SPIN POLARIZATION AT A CONSTANT CONCENTRATION OF CARRIERS

Consider the solution to the general equations obtained in the preceding section for the particular case of time-independent transition ν_{tr} and relaxation Λ_2 rates and muon spin precession in a transverse field. Precisely such precession has been studied experimentally the most thoroughly [2–11].

For the first time, the equations that described the behavior of transverse polarization $P_\perp(t)$ were obtained in [7],

$$P_\perp(t) = P_1\exp(-(\nu_{tr} + \Lambda_2)t) \times \cos(\omega_p t + \varphi_1) + P_2\cos(\omega_d t + \varphi_2). \quad (17)$$

The damped P_1 and nonrelaxing P_2 amplitude components of the μSR signal are related to the initial polarization in the diamagnetic fraction $P_0 \equiv P_\perp(d, 0)$, the transition rate ν_{tr}, and the frequency difference $\delta\omega = \omega_p - \omega_d$ as

$$P_1 = (1 - P_0)\sqrt{\frac{\Lambda_2^2 + \delta\omega^2}{(\nu_{tr} + \Lambda_2)^2 + \delta\omega^2}},$$
$$P_2 = \sqrt{\frac{(P_0\Lambda_2 + \nu_{tr})^2 + P_0^2\delta\omega^2}{(\nu_{tr} + \Lambda_2)^2 + \delta\omega^2}}. \quad (18)$$

The phases of the corresponding components are determined by the equations

$$\tan\varphi_1 = \frac{-\nu_{tr}\delta\omega}{\nu_{tr}\Lambda_2 + \Lambda_2^2 + \delta\omega^2},$$
$$\tan\varphi_2 = \frac{(1 - P_0)\nu_{tr}\delta\omega}{(\nu_{tr}\Lambda_2 + \Lambda_2^2 + \delta\omega^2)P_0 + \nu_{tr}\Lambda_2 + \nu_{tr}^2}. \quad (19)$$

According to (9), the paramagnetic shift $\delta\omega$ only depends on temperature T and magnetic field B (because the ω_d frequency is proportional to the magnetic field). Equation (9) also shows that the relaxation rate of the muon moment Λ_2 is determined by the relaxation rate of the electron shell moment ν_e and the precession frequency of the electronic moment ω_e. Electronic moment relaxation is caused by interaction with phonons CT^q and exchange interactions of two types,

$$\nu_e(T) = CT^q + \nu_0 + \sigma_{ex}v_T(T)n(T). \quad (20)$$

The parameters of the first term, which describes the phonon relaxation mechanism, $C \approx 7 \times 10^6$ s^{-1} and $q \approx 3$, follow from the experimental data obtained in [10]. The $q \approx 3$ estimate given in [10] is only valid at temperatures below 30 K. Already at about 20 K, the electronic moment relaxation rate becomes comparable to the electron precession frequency ω_e. It follows that, if the law T^3 holds, the paramagnetic shift at the higher temperatures is unobservable in principle because $\nu_e > \omega_e$ [1].

The second term ν_0 describes the exchange interaction of the acceptor center with neighboring nonionized impurities. It is determined by the impurity concentra-

tion and weakly depends on temperature, as was substantiated experimentally [8–11] in the temperature range $5 < T < 20$ K. At impurity concentrations below 10^{17} cm^{-3}, a noticeable temperature dependence of the rate of polarization damping is evidence that the contribution of ν_0 to electronic moment relaxation is small. The ν_0 value at a high impurity concentration can be estimated from the experimental temperature dependence of the rate of muon polarization damping [8–11]. The results obtained in [8–11] show that the rate ν_0 of the interaction of the muon acceptor center with an impurity acceptor center is on the order of 10^{11} s^{-1}, and the rate of the interaction of the muon acceptor center with an impurity donor center is two orders of magnitude lower. It follows that the rate of the exchange interaction of the muon acceptor center with a donor impurity should be $\nu_0 \sim 10^9$ s^{-1} at $N_d \sim 10^{18}$ cm^{-3}.

The last term in (20) is responsible for the exchange scattering of free carriers by the electron shell of the acceptor center and is determined by the concentration of free electrons $n(T)$, the cross section of exchange scattering σ_{ex}, and the mean thermal velocity of carriers $v_T(T)$ [cm/s] $\approx 10^5\sqrt{T\,[\mathrm{K}]}$. The cross section of exchange scattering by the acceptor center at 4–10 K can be estimated using phase theory. In the limit of low collision energies $ka_B \longrightarrow 0$ for the model of a hydrogen-like atom, this theory gives [25]

$$\sigma_{ex}(T) = \frac{\pi}{k^2}\sin^2(\delta_0 - \delta_1),$$

where $k = mv_T/\hbar$ is the wave vector modulus; δ_0 and δ_1 are the singlet and triplet scattering phases, respectively; and a_B is the characteristic size of the acceptor center. The phases tend to π linearly in k as $ka_B \longrightarrow 0$; therefore, $\delta_1 - \delta_0 = 3.86ka_B$ [25], and $\sigma_{ex} \approx 15\pi a_B^2$. On the assumption that $a_B \approx a_0\varepsilon m_e/m_{\mathrm{eff}} \approx 30$ Å [24] (here, $\varepsilon = 11$ is the dielectric constant of silicon, a_0 is the Bohr radius, $m_{\mathrm{eff}} = 0.22m_e$ is the effective mass of the electron in silicon, and m_e is the mass of the free electron), we obtain $\sigma_{ex} \approx 4 \times 10^{-12}$ cm^2; this value was used in the calculations. The above estimate is valid for the exchange scattering of electrons by a donor center. The σ_{ex} value for the scattering of electrons by the acceptor center calculated taking into account the complete wavefunction of the acceptor center may, however, differ from this estimate by as much as an order of magnitude. However, importantly, σ_{ex} is constant in the limit of low temperatures, and this value should therefore be adjusted for correctly analyzing experiments.

In pure samples and at low concentrations of doping impurities, the contribution of the last two terms is zero. The influence of exchange scattering was observed experimentally in [11]; it manifested itself by a deviation of the temperature dependence of muon moment relaxation rate Λ_2 from the predicted law.

The transition from the paramagnetic ($_\mu Al^0$) to diamagnetic ($_\mu Al^-$) acceptor center state involves conduction electron capture and is characterized by the transition rate

$$\nu_{\mathrm{tr}}(T) = \kappa_{\mathrm{tr}} n(T). \tag{21}$$

The $\kappa_{\mathrm{tr}} = \sigma_{\mathrm{tr}} v_T(T)$ value is called the capture coefficient; it is virtually independent of temperature at $T < 100$ K [26]. Unambiguous and noncontradictory experimental data on this value for the Al impurity center in silicon are lacking. It was estimated at 2×10^{-13} cm^3 s^{-1} from the μSR data obtained in [8] in the temperature range $4.5 < T < 10$ K, which corresponded to capture cross sections $\sigma_{\mathrm{tr}} < 10^{-20}$–$10^{-19}$ cm^2. The correctness of this estimate is questionable, because the σ_{tr} value is then substantially smaller than the atomic cross section. In this work, we use an alternative estimate of κ_{tr} suggested in [27], namely, $\kappa_{\mathrm{tr}} = 10^{-7}$ cm^3/s at $5 < T < 50$ K, which corresponds to the cross section $\sigma_{\mathrm{tr}} \sim 10^{-14}$ cm^2.

The equilibrium concentration of electrons $n(T) = n_T(T)$ in nondegenerate n-type semiconductors can very accurately be approximated as [27]

$$n_T(T) = \frac{2N_d}{1 + \sqrt{1 + 8\dfrac{N_d}{N_C(T)}\exp\left(\dfrac{\varepsilon_d}{kT}\right)}}, \tag{22}$$

where

$$N_C(T) = \frac{1}{\sqrt{2}}\left(\frac{\sqrt{\pi m_{\mathrm{eff}} k_B T}}{\pi\hbar}\right)^3,$$

ε_d is the energy of impurity ionization (39 meV for antimony and 45 meV for phosphorus), k_B is the Boltzmann constant, and N_d is the concentration of the donor impurity.

A nonrelaxing component was experimentally observed in silicon in [6, 7] for a sample with the concentration of donors (antimony) $N_d = 2 \times 10^{18}$ cm^{-3}. The authors of [6, 7] analyzed their experimental data on the assumption that $P_0 = 0$. However, $P_0 = 0.1$ corresponded to this concentration of donors. This is in agreement with the observation that the amplitudes of the relaxing and nonrelaxing polarization components (symbols in Fig. 1) asymptotically tend to 0.9 and 0.1, respectively, at low temperatures. According to calculations, the incorrect assumption on P_0 leads to errors of about 10% in Λ_2 and ν_{tr}. In addition, the authors of [6, 7] analyzed the experimental $P_\perp(t)$ curves using $\delta\omega$ as an adjustment parameter to determine the hyperfine interaction constant for $N_d = 2 \times 10^{18}$ cm^{-3}. The accuracy of

the determination was, however, low: $\delta\omega/\omega_\mu$ was of $(5–10)\times10^{-3}$ in the temperature range 4–20 K. According to (9), $\delta\omega$ is independent of the impurity concentration N_d, and the A values estimated in [6, 7] are not at variance with the more accurate data obtained at $N_d < 10^{16}$ cm^{-3}. It is therefore more reasonable to perform data processing using the temperature dependence of $\delta\omega$ averaged over a large number of experiments in pure and weakly doped silicon [28]. For this reason, we performed calculations by (9) to estimate $\delta\omega$ at A = 25 MHz [28].

The temperature dependences of the polarization amplitudes P_1 and P_2 and precession phases φ_1 and φ_2 constructed according to (18) and (19) with the parameters were set equal to the values discussed above are shown in Fig. 1. The ν_0 value was set at 10^9 s^{-1}. The slopes of the $P_1(T)$ and $P_2(T)$ curves strongly depend on $\nu_{tr}(T)$ at temperatures from 4 to 10 K. It follows from Fig. 1 that, at $T \approx 4$ K, when the concentration of equilibrium electrons $n_T(T)$ is negligibly small, $\nu_{tr}(T)$ is nevertheless not too small; otherwise, $P_1(T)$ and $P_2(T)$ would weakly depend on temperature. It follows that nonequilibrium electrons are present at the track endpoint and the concentration n_e is therefore much higher than the equilibrium value.

To fit the calculated curves to the experimental data [6, 7] in the temperature range 4–10 K, we must put $n_e\kappa_{tr} \approx 2\times10^6$ s^{-1}, which, considering the $\kappa_{tr} \approx 10^{-7}$ cm^3/s value accepted above, gives $n_e \approx 2\times10^{13}$ cm^{-3}. The curve shapes at higher temperatures of 10 to 15 K are largely determined by changes in the concentration of equilibrium carriers $n_T(T)$, the total concentration of electrons being $n(T) = n_e + n_T(T)$.

Experimental data processing was performed in [6, 7] not quite correctly, even on the assumption that the ν_e and ν_{tr} rates were constant in time. The condition $P_1 + P_2 = 1$ used in [6, 7] holds to a high accuracy and does not give a noticeable error (Fig. 1). However, the ignoring of phase φ_1 and φ_2 changes at $T > 15$ K, where these changes are substantial, leads to incorrect interpretation of the temperature dependences of fraction amplitudes and the rates ν_{tr} and Λ_2. Treatment of the experimental data from [6, 7] by (18) and (19) in the temperature range 15–20 K should ensure coincidence with the theoretical curves shown in Fig. 1.

The $\nu_{tr}(T)$ and $\Lambda_2(T)$ dependences, calculated and reproduced in [6, 7], are shown in Fig. 2. For Λ_2, our results qualitatively coincide with those obtained in [6, 7]. As to the rate of capture ν_{tr}, it was found to decrease at high temperatures in [6, 7], which had no physical interpretation. Conversely, the theoretical dependence predicts a sharp increase in this value. To avoid such inaccuracies in reproducing the ν_{tr} and Λ_2 parameters (in particular, at temperatures about 15 K, where $\nu_{tr} \sim \Lambda_2 \sim \delta\omega$), we must take into account the relation between the adjustment parameters that follows from (9), (18), and (19) and use the A value obtained for pure or weakly doped samples.

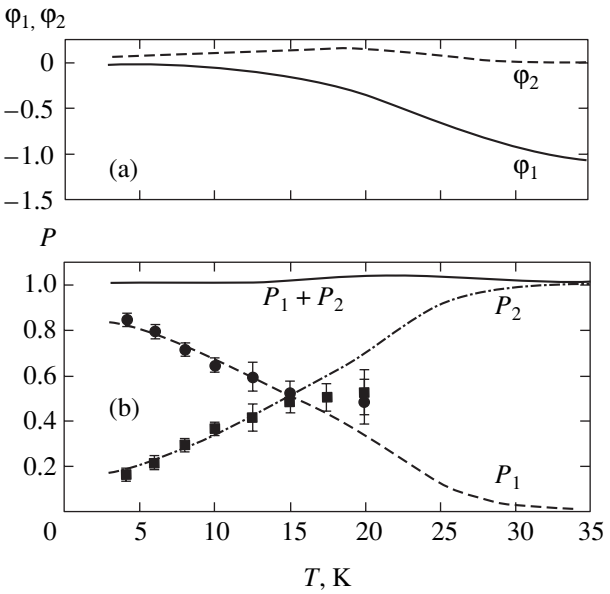


Fig. 1. (a) Calculated temperature dependences of the phases of the damped φ_1 and unrelaxed φ_2 components and (b) calculated temperature dependences of the polarization amplitudes of the damped P_1 and nonrelaxed P_2 components. The calculations were performed for n-silicon with antimony concentration [Sb] = 2×10^{18} cm^{-3} at B = 1 kG; experimental data are shown by symbols [6, 7].

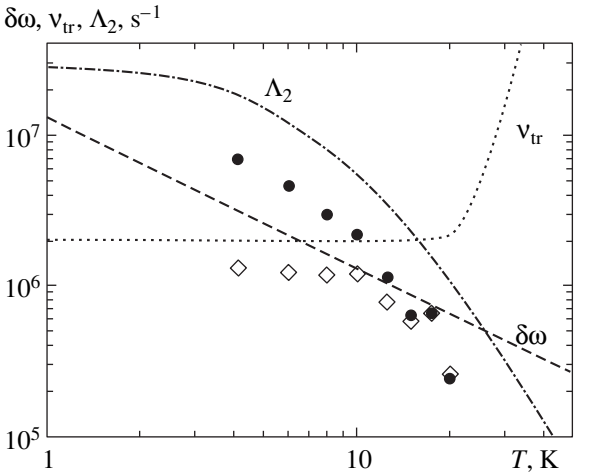


Fig. 2. Temperature dependences of paramagnetic shift $\delta\omega$, relaxation rate Λ_2 (solid circles are experimental data from [6, 7]), and the rate of transition to the diamagnetic state ν_{tr} (rhombi are the experimental data from [6, 7]); conditions are the same as in Fig. 1.

4. RECOMBINATION AT THE TRACK ENDPOINT

Nonequilibrium electrons at the muon track endpoint in samples with a substantial concentration of

impurities appear as a result of medium ionization by the muon and impurity ionization by secondary and cascade Auger electrons whose energy is much higher than the impurity ionization potential. The cross sections of impurity ionization by electrons with energies on the order of 1 eV may exceed 10^{-10} cm^2 [27]. For this reason, the initial concentration of electrons and ionized impurities at the track endpoint can be on the order of the impurity concentration [1].

The recombination of electrons with ionized impurities rapidly decreases their concentration. This process is characterized by the recombination constant [29]

$$K(T)\ [\mathrm{cm}^3/\mathrm{s}] = \frac{3.3\times 10^{-3}}{(4.9 + T\ [\mathrm{K}])^2\sqrt{T\ [\mathrm{K}]}}. \qquad (23)$$

Estimates show that, at such a recombination constant value, the concentration of electrons should be lower than the $n_c \approx 2\times 10^{13}$ cm^{-3} concentration obtained above by fitting the calculation results to experimental data at times longer than 10^{-8} s, which are observed with confidence in μSR experiments. This discrepancy arises because (23) is valid for electron capture by a solitary Coulomb center if the screening of the charge of this center is insignificant; that is, if $r_T \ll r_D$, where $r_D = \sqrt{k_B T/4\pi e^2 n}$ is the Debye radius and r_T is the Thomson radius. The Thomson radius is determined by the condition $U(r_T) = k_B T$. For an unscreened Coulomb center, the "classical" Thomson radius is $r_{Tc} = e^2/k_B T$.

At $T = 6$ K and $n_c \approx 2\times 10^{13}$ cm^{-3}, we have $r_D = 7\times 10^{-6}$ cm, $r_{Tc} = 3\times 10^{-4}$ cm, and $r_{Tc} > r_D$. Accordingly, Eq. (23) from [29] based on the Thomson model requires refinement.

An electron produced during muon stopping loses its initial energy in time $\tau_T\,[\mathrm{s}] = 3.9\times 10^{-10}\sqrt{T[\mathrm{K}]}$ even in one-phonon processes only [29]. For this reason, all electrons are thermalized and a substantial fraction of them are recombined with impurity ions at times $t > 10^{-9}$ s. In the Thomson model, the capture (recombination) cross section is proportional to r_T^3/l_ε, where $l_\varepsilon = 7.5\times 10^{-4}$ cm is the energy loss length [29]. The capture cross section can be corrected by describing the attractive potential as the screened Coulomb potential $U(r) = \exp(-r/r_s)e^2/r$, where r_s is the screening radius. Accordingly, the generally accepted recombination constant (23) should be multiplied by the coefficient $\zeta(n(t, T), T)$,

$$\zeta(n, T) = \left(\frac{r_T(n, T)}{r_{Tc}(n, T)}\right)^3. \qquad (24)$$

The screening radius r_s for finding r_T is set equal to the Debye radius, $r_s = r_D$, as long as $r_D(n, T) > n^{-1/3}$. If the Debye radius is smaller than the distance between charged particles, this description of interaction screening becomes incorrect. There is no consistent recombination theory for this situation. For this reason, when the condition of Debye theory applicability is violated and $e^2 n^{1/3} > T$, estimative calculations are performed with the screening radius set equal to half the mean distance between charged particles, $r_s = 0.5n^{-1/3}$, which virtually coincides with the radius of the Seitz–Wigner cell $(4/3\pi n)^{-1/3}$.

Correction (24) qualitatively changes the temperature dependence of the recombination constant. If $r_{Tc} > r_D$, r_T reaches the order of the screening radius r_D, which is proportional to $\sqrt{T}$. Accordingly, in the plasma temperature and concentration region where $r_{Tc} > r_D$, the recombination constant $K(T)$ increases as temperature grows. In the region where $r_T < r_D$, this increase changes for habitual lowering, because $r_{Tc} \propto 1/T$.

Recombination is described by the equation

$$\frac{dn(t)}{dt} = -\zeta(n(t), T)K(T)(n(t) - n_T(T))^2, \qquad (25)$$

which can be solved with respect to the time variable,

$$t(n_0, n) = \frac{1}{K(T)}\int_n^{n_0} \frac{dn'}{\zeta(n', T)(n' - n_T(T))^2}. \qquad (26)$$

The parametric representation $t = t(n_0, n)$ allows calculations to be reduced to the calculation of integrals and all dependences to be obtained in the form of quadratures.

The time dependences of the concentration of electrons calculated in the approximation described above at $T = 6$ K and various initial electron concentrations n_0 are shown in Fig. 3. We see from this figure that the concentration of electrons is almost independent of n_0 in the range 10^{14}–10^{18} cm^{-3} at times on the order of 10^{-7} s. For comparison, line *4* (Fig. 3) was calculated ignoring screening; it corresponds to the asymptotic behavior of the dependences obtained in more correct calculations. The concentration of electrons cannot be considered constant at 6 K at times of 10^{-8} to 10^{-6} s, which are usual for μSR measurements.

In the calculations described in Section 3, the concentration of nonequilibrium electrons by a time of 10^{-8}–10^{-7} s was set equal to $n_c \approx 2\times 10^{13}$ cm^{-3}. This concentration was selected from the requirement on the coincidence of the slopes of the calculated and experimental temperature dependences of polarization at $T <$ 10 K. As follows from Fig. 3, the $n_c \approx 2\times 10^{13}$ cm^{-3} value falls within the range of concentrations that in reality exist at times of 10^{-8}–10^{-7} s. This explains the qualitative agreement between the polarization curves calculated on the assumption that n is constant and the measurement results [7] (Fig. 1). The use of the model with a constant concentration of electrons for con-

structing the temperature dependences $\nu_{tr}(T)$ and $\Lambda_2(T)$ in the temperature range 4–10 K, however, makes no sense, because these values are more sensitive than polarization to the time-dependent n value. At $T > 10$ K, when the equilibrium concentration of electrons determined by (22) becomes predominant, n ceases to depend on time, and the temperature dependences shown in Figs. 1 and 2 should then correctly describe the experimental results.

5. THE INFLUENCE OF THE NONSTATIONARY CONCENTRATION OF NONEQUILIBRIUM CARRIERS ON THE BEHAVIOR OF MUON SPIN POLARIZATION

As previously, the total concentration of electrons is represented as the sum of the concentrations of equilibrium and nonequilibrium electrons $n(t, T) = n_T(T) + n_c(t, T)$. As electrons are fairly rapidly thermalized and their mean energy ceases to depend on time, the p–d transition constant of the acceptor center $\kappa_{tr} = v_T\sigma_{tr}$ can be considered time independent at $t > 10^{-9}$ s. At energies higher than thermal, the cross section of electron capture by a neutral center sharply decreases. For this reason, the recombination of electrons at times shorter than 10^{-9} s has low probability and can be ignored. The time $t_{pd} = 10^{-9}$ s will be taken as the initial time for the p–d transition. By this time, nonequilibrium carriers are already thermalized, and their equilibrium concentration is in the range 10^{13}–10^{16} cm^{-3} no matter which approximation of those described in the preceding section is used (Fig. 3). Precisely this concentration n_0 will be considered "initial" for all processes that occur at $t > 10^{-9}$ s.

The probability $w(p, t)$ that the acceptor center will remain paramagnetic by time t can be obtained by substituting (21) into (11). According to (26), t can be represented in terms of $n(t, T)$. Using this representation, we obtain

$$w(p, t(n_0, n)) = w_p(n_0, n) = \exp\left\{-\frac{\kappa_{tr}}{K(T)}\int_n^{n_0}\frac{n'dn'}{\zeta(n', T)(n' - n_T(T))^2}\right\}. \quad (27)$$

Equation (27) is valid if $n(t, T) > n_T(T)$. At $t > t(n_0, n_T(T))$, the concentration of free electrons reaches its equilibrium thermal value and ceases to change. Further polarization behavior is described by the equations given in Section 3.

If screening is included, (27) requires numerical integration. For comparison, we give an analytic equation that can be obtained if screening is ignored and it is assumed that $n(t, T) \gg n_T(T)$,

$$w(p, t, T) = \exp\left(-\kappa_{tr}(T)\left\{n_T(T)t + \frac{\ln(n_0K(T)t + 1)}{K(T)}\right\}\right). \quad (28)$$

The $w(p, t, T)$ dependence should replace the exponential multiplier in (13) that takes into account the probability of transitions into the diamagnetic state.

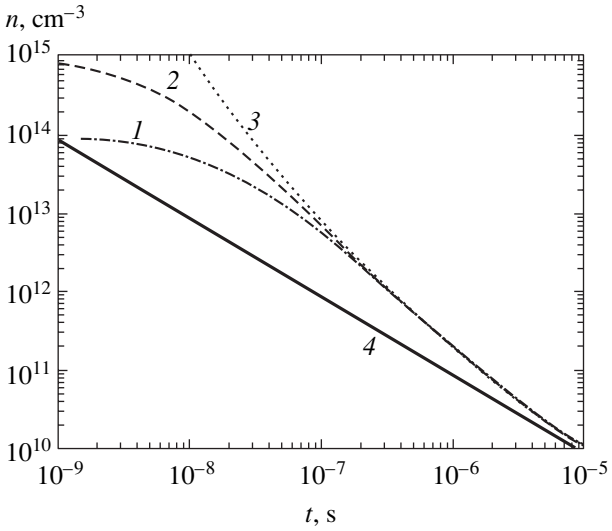


Fig. 3. Time dependences of the concentration of nonequilibrium charge carriers at the muon track endpoint: (*1*)–(*3*) calculations by (26) for the initial electron concentrations $n_0 = 10^{14}$, 10^{15}, and 10^{18} cm^{-3}, respectively, and (*4*) calculations without taking into account Debye screening with $\zeta \equiv 1$; $T = 6$ K.

The $w(p, t, T)$ time dependences obtained by (27) at $T = 6$ K for various initial concentrations n_0 are shown in Fig. 4. These curves are substantially different from the exponential function used in [6, 7] to approximate the experimental data.

These dependences can be used to obtain the upper estimate for n_0 under experimental conditions [6]. As neither "lost" polarization nor substantial (in excess of 10%) changes in the amplitude of paramagnetic component precession were observed at $T = 6$ K at times shorter than 10^{-8} s, we obtain $n_0 < 10^{14}$ cm^{-3}.

The $\Lambda_2(n)$ value only depends on time via the $n(t, T)$ variable concentration of nonequilibrium carriers. We can therefore use (26) to write the corresponding multiplier in Eq. (13) for paramagnetic fraction polarization in the form convenient for calculations,

$$F(p, t(n_0, n)) = F_p(n_0, n) = \exp\left\{-\frac{1}{K(T)}\int_n^{n_0}\frac{\Lambda_2(n', T)dn'}{\zeta(n', T)(n' - n_T(T))^2}\right\}. \quad (29)$$

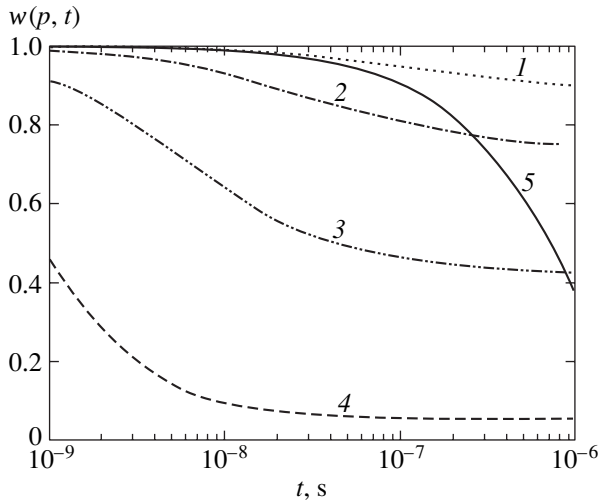


Fig. 4. Probability $w(p, t)$ for an acceptor center to remain neutral: (*1*)–(*4*) calculations by (27) including screening for $n_0 = 10^{13}$, 10^{14}, 10^{15}, and 10^{16} cm^{-3}, respectively, and (*5*) exponential approximation [7]; $T = 6$ K and $B = 1$kG.

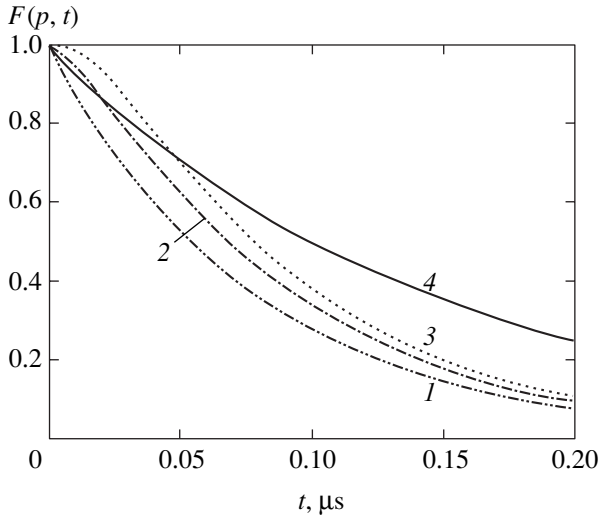


Fig. 5. Time behavior of the $F(p, t)$ function in a neutral acceptor center: (*1*)–(*3*) calculations by (29) including screening at $n_0 = 10^{13}$, 10^{15}, and 10^{18} cm^{-3}, respectively, and (*4*) exponential dependence with $\Lambda_2 = 7 \times 10^6$ s^{-1} [7]; $T = 6$ K and $B = 1$ kG.

According to (9), the dependence of Λ_2 on n and, accordingly, time is contained in the frequency $\nu_e(t, T)$ described by (20).

The polarization "amplitudes" $F(p, t)$ in the paramagnetic fraction calculated by (29) using various n_0 values are shown in Fig. 5 for $B = 1$ kG and $T = 6$ K. The other parameters were the same as used in constructing the curves shown in Figs. 3 and 4. The difference in the behaviors of the curves that describe the probability for the acceptor center to remain in the paramagnetic state (Fig. 4) and the polarization "amplitude" in the paramagnetic fraction (Fig. 5) is noteworthy. This difference can be used to separate the contributions of the *p*–*d* transition and muon spin relaxation processes to the time dependence of the transverse polarization component, which precesses at frequency ω_p.

To separate the electron capture and muon spin relaxation contributions and thereby more accurately reproduce cross sections from μSR experimental results, we must perform measurements in substantially different magnetic fields, because the rate of polarization relaxation Λ_2 depends on B [see (9)]. At times of 10^{-7} s, the polarization amplitude in "zero" field differs approximately twofold from its value at $B = 1$ kG. If zero field is inconvenient for measurements, the results obtained in 500 and 1500 G fields can be compared; these results should differ by approximately 30%. This accuracy of determining amplitudes is quite attainable in modern μSR experiments.

The diamagnetic fraction contribution to the precession signal at the ω_p frequency was found to be small. For comparison, we give the results of calculations of the behavior of polarization in the diamagnetic fraction by the formula

$$P_\perp(d, t(n_0, n)) = P_\perp(d, 0)\cos(\omega_d t(n_0, n)) + P_\perp(p, 0)\{\mathrm{Re}[J(n_0, n)]\cos(\omega_d t(n_0, n)) + \mathrm{Im}[J(n_0, n)]\sin(\omega_d t(n_0, n))\}, \quad (30)$$

where

$$J(n_0, n) = \int_n^{n_0} \frac{\kappa_{tr} n' \exp(i\omega_d t(n_0, n'))}{(n' - n_T(T))^2 \zeta(n', T) K(T)} \times w_p(n_0, n') F_p(n_0, n') \cos(\omega_p t(n_0, n')) dn'.$$

This formula follows from (14) after the replacement of $\cos(\omega_d(t - t'))$ by the product of cosines and sines. The time behavior of the $\mathrm{Re}[J(n_0, t(n_0, n))]$ and $\mathrm{Im}[J(n_0, t(n_0, n))]$ coefficients determined using (28) and (29) is shown in Fig. 6. Note that, under the conditions specified, the contribution of the second term in (30) is inessential at concentrations $n_0 < 10^{14}$ cm^{-3} estimated above.

Comparison of the time behavior of the summed polarization in a silicon sample with an antimony concentration $N_d = 2 \times 10^{18}$ cm^{-3} calculated by the equations given above at $T = 6$ K with the experimental values and the approximating curve from [7] shows that the two curves coincide with each other and with experimental data to within measurement errors. It has, however, been demonstrated above that neither an exponential time dependence of the probability of the occurrence of an acceptor center in the paramagnetic fraction nor exponential relaxation (as suggested in [7] in constructing the approximating curve) are observed in the sample under consideration. The coincidence of the

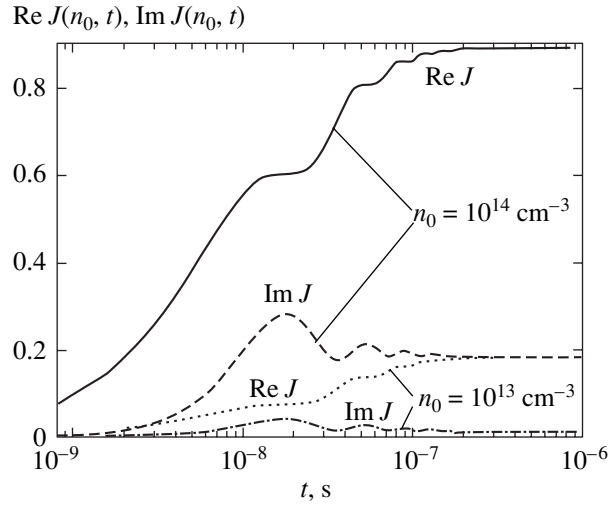


Fig. 6. Time dependences of the real and imaginary parts of $J(n_0, t)$ at $n_0 = 10^{13}$ and 10^{14} cm^{-3}, $T = 6$ K, and $B = 1$ kG.

approximating curve with experimental values in [7] was obtained by using a large amount adjustment parameters. The $\nu_{tr}(T)$ and $\Lambda_2(T)$ values then become physically meaningful only at temperatures above 15 K, when the equilibrium concentration of electrons $n_T(T) \sim 10^{11}$ cm^{-3} is comparable to their nonequilibrium concentration (see Fig. 3) and the concentration of electrons ceases to depend on time. In this respect, our interpretation of experimental data based on a variable concentration of nonequilibrium carriers appears to be better grounded because all its elements have real physical meaning.

To summarize, including the concentration of nonequilibrium carriers in a muon track plasma and its changes in time makes it possible to approximate experimental data on muon spin polarization in doped silicon samples at low temperatures (below 15 K) and use these data to obtain approximate estimates of the κ_{tr}, σ_{ex}, and n_0 values.

6. CONCLUSIONS

The mere possibility of observing negative muon spin precession at the muon frequency in doped nondegenerate semiconductors at low temperatures can be explained by the effects of track free carriers, which are responsible for the transition from the paramagnetic into diamagnetic fraction.

The generalization of the Vangsness–Bloch equations that we suggest allows the behavior of muon spin polarization under the conditions of a variable density of electrons at the muon track endpoint to be described. An analysis based on these equations gives a more correct interpretation of the available experimental results of μSR studies of semiconductors performed using negative muons.

In nondegenerate semiconductors, the μSR method allows the cross section of electron capture by an acceptor center and a solitary charged Coulomb center to be reproduced in the order of magnitude and information about the cross sections of exchange scattering of electrons and holes by impurity centers at very low energies inaccessible to direct measurements to be obtained.

In order to obtain useful information on kinetic processes in the region of muon localization by the μSR method, the approximating function for polarization should be constructed with the use of parameters that are indeed constant in time. These are the constant of electron capture by an acceptor center κ_{tr} and spin exchange scattering cross sections σ_{ex} rather than the transition ν_{tr} and polarization relaxation Λ_2 rates. The relaxation rate of the electronic moment of an acceptor center in interactions with phonons should be characterized by parameters obtained from experiments with samples containing lower impurity concentrations.

Nor does the use of the value of the paramagnetic shift as an adjustment parameter for doped samples make sense, because, at impurity concentrations on the order of 10^{18} cm^{-3}, we observe paramagnetic shifts approximately equal to that in pure samples. There is no sufficient reason to believe that the hyperfine interaction constant A and, accordingly, the paramagnetic shift value strongly change in a nondegenerate semiconductor as the concentration of impurities increases. The A value can change in a strongly internally strained sample; the role played by internal strain requires additional inquiries.

If the Debye radius in a solid-state muon track plasma is larger than the mean distance between charged particles, measurements of amplitude damping for muon spin polarization and the ratio between the polarization amplitudes of different fractions can be used to reproduce the concentration dependence of the recombination "constant" on impurity centers in a solid-state semiconductor plasma. If the Debye radius in such a plasma is smaller than the Thomson radius, we must construct a new theory of electron ion recombination. In this work, we qualitatively estimated changes in the recombination constant in such a situation. A consistent theory has not been suggested thus far, but experimental studies of this problem by the μ^-SR method can be useful for its development.

In this work, we restricted consideration to the results of μSR experiments for nondegenerate doped silicon, because such samples have been studied the most thoroughly by the μSR method. We must also gain a deeper understanding of the first experiments with heavily doped silicon samples in the region of the transition to metallic conductivity [10, 11] to correctly interpret them.

ACKNOWLEDGMENTS

This work was financially supported by the Russian Foundation for Basic Research (project nos. 02-02-16370-a and 02-02-16881-a) and the Ministry of Education of the Russian Federation (Universities of Russia Program, project no. UR.01.01.015).

REFERENCES

1. V. N. Gorelkin, T. N. Mamedov, and D. V. Rubtsov, Hyperfine Interact. C **1**, 191 (1996).
2. V. N. Gorelkin, V. G. Grebennik, K. I. Gritsaĭ, *et al.*, Yad. Fiz. **56** (10), 29 (1993) [Phys. At. Nucl. **56**, 1316 (1993)].
3. T. N. Mamedov, V. N. Duginov, V. G. Grebennik, *et al.*, Hyperfine Interact. **86**, 717 (1994).
4. T. N. Mamedov, V. N. Gorelkin, V. G. Grebennik, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **63**, 539 (1996) [JETP Lett. **63**, 566 (1996)].
5. T. N. Mamedov, I. L. Chaplygin, V. N. Duginov, *et al.*, Hyperfine Interact. **105**, 345 (1997).
6. T. N. Mamedov, V. N. Duginov, D. Gerlach, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **68**, 61 (1998) [JETP Lett. **68**, 64 (1998)].
7. T. N. Mamedov, I. L. Chaplygin, V. N. Duginov, *et al.*, J. Phys.: Condens. Matter **11**, 2849 (1999).
8. T. N. Mamedov, D. G. Andrianov, D. Gerlach, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 637 (2000) [JETP Lett. **71**, 438 (2000)].
9. T. N. Mamedov, K. I. Gritsaj, A. V. Stoikov, *et al.*, Physica B (Amsterdam) **289–290**, 574 (2000).
10. T. N. Mamedov, D. G. Andrianov, D. Gerlach, *et al.*, Zh. Éksp. Teor. Fiz. **119**, 1159 (2001) [JETP **92**, 1004 (2001)].
11. T. N. Mamedov, D. G. Andrianov, D. Gerlach, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **73**, 759 (2001) [JETP Lett. **73**, 674 (2001)].
12. N. Korst, Teor. Mat. Fiz. **6**, 265 (1971).
13. I. V. Aleksandrov, *Theory of Magnetic Relaxation* (Nauka, Moscow, 1975) [in Russian].
14. V. G. Nosov and I. V. Yakovleva, Zh. Éksp. Teor. Fiz. **43**, 1751 (1962) [Sov. Phys. JETP **16**, 1236 (1963)].
15. I. G. Ivanter and V. P. Smilga, Zh. Éksp. Teor. Fiz. **54**, 559 (1968) [Sov. Phys. JETP **27**, 301 (1968)].
16. I. G. Ivanter and V. P. Smilga, Zh. Éksp. Teor. Fiz. **55**, 1521 (1968) [Sov. Phys. JETP **28**, 796 (1969)].
17. I. G. Ivanter and V. P. Smilga, Zh. Éksp. Teor. Fiz. **60**, 1985 (1971) [Sov. Phys. JETP **33**, 1070 (1971)].
18. I. G. Ivanter and V. P. Smilga, Zh. Éksp. Teor. Fiz. **61**, 2176 (1971) [Sov. Phys. JETP **34**, 1167 (1972)].
19. V. P. Smilga and Yu. M. Belousov, *The Muon Method in Science* (Nauka, Moscow, 1991; Nova Sci., New York, 1994).
20. A. S. Baturin and V. N. Gorelkin, Physica B (Amsterdam) **289–290**, 578 (2000).
21. V. N. Gorelkin, T. N. Mamedov, and A. S. Baturin, Physica B (Amsterdam) **289–290**, 585 (2000).
22. V. A. Karasyuk, M. W. Thewald, S. An, *et al.*, Phys. Rev. B **54**, 10543 (1996).
23. A. Abragam and B. Bleaney, *Electron Paramagnetic Resonance of Transition Ions* (Clarendon Press, Oxford, 1970; Mir, Moscow, 1973).
24. A. M. Stoneham, *Theory of Defects in Solids: the Electronic Structure of Defects in Insulators and Semiconductors* (Clarendon Press, Oxford, 1975; Mir, Moscow, 1978).
25. N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions*, 3rd ed. (Clarendon Press, Oxford, 1965; Mir, Moscow, 1969).
26. 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).
27. A. G. Milnes, *Deep Impurities in Semiconductors* (Wiley, New York, 1973; Mir, Moscow, 1977).
28. T. N. Mamedov, A. V. Stoĭkov, and V. N. Gorelkin, Fiz. Élem. Chastits At. Yadra **33**, 1005 (2002) [Phys. Part. Nucl. **33**, 519 (2002)].
29. V. N. Abakumov, V. I. Perel', and I. N. Yasievich, Fiz. Tekh. Poluprovodn. (Leningrad) **12**, 3 (1978) [Sov. Phys. Semicond. **12**, 1 (1978)].

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 2, 2004, pp. 363–369.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 2, 2004, pp. 407–414.

SOLIDS
Electronic Properties

Phase Separation in $La_{2-x}A_xCoMnO_6$ (A = Ca and Sr) Perovskites

I. O. Troyanchuk[a], A. P. Sazonov[a], H. Szymczak[b], D. M. Többens[c], and H. Gamari-Seale[d]

[a]*Institute of Solid State and Semiconductor Physics, National Academy of Sciences of Belarus, Minsk, 220072 Belarus*
e-mail: troyan@ifttp.bas-net.by
[b]*Institute of Physics, Polish Academy of Sciences, 02-668, Warsaw, Poland*
[c]*Berlin Neutron Scattering Center, Hahn-Meitner-Institut, 14109, Berlin, Germany*
[d]*Institute of Material Science, National Center of Scientific Research "Demokritos," 15343, Athens, Greece*
Received December 18, 2003

Abstract—Magnetic and neutron-diffraction studies of La_2CoMnO_6 doped with strontium and calcium ions are carried out. Stoichiometric and calcium-containing samples are orthorhombically distorted, whereas the oxidation and doping with strontium ions stabilizes a rhombohedral structure. Neutron-diffraction studies show that La_2CoMnO_6 is a ferromagnet, whereas a Ca-substituted compound does not exhibit long-range magnetic order. The whole set of data suggest that these compounds are inhomogeneous magnets that consist of clusters in which Co^{2+} and Mn^{4+} ions are ordered similar to ions in NaCl and of regions in which there is no such ordering. These regions exhibit different magnetic properties. Phase separation is discussed in a model of solid solutions with intrinsic chemical inhomogeneity on the nanometer scale. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Compounds based on lanthanum manganite show a wide variety of interesting properties due to the interaction between orbital, charge, and magnetic degrees of freedom. In spite of a large number of studies, the nature of many phenomena is still a matter of dispute. Magnetic phase separation in lanthanum manganites is one of such phenomena. This phenomenon was observed at micro- and macroscopic levels in many solid solutions of the types $Ln_{1-x}Ca_xMnO_3$ and $Ln_{1-x}Sr_xMnO_3$ (Ln = La and Pr). An especially clear example of this phenomenon was observed near antiferromagnet–ferromagnet concentration phase transitions, for example, in $La_{0.95}Sr_{0.05}MnO_3$ [1] or $La_{0.5}Ca_{0.5}MnO_3$ [2] compounds. At present, it is believed that magnetic inhomogeneity plays an important role in the formation of the magnetoresistance effect [3]. Note that magnetic inhomogeneity was also observed in $La_{1-x}Sr_xCoO_3$ cobaltites [4], as well as in La_2CoMnO_6-type perovskites in which cobalt and manganese ions are partially ordered [5]. In [6], the authors observed a peak on the temperature dependence of the dynamic magnetic susceptibility below the Curie temperature (T_C = 230 K) and interpreted it in terms of a magnetic two-phase state. According to [7, 8], the magnetic two-phase state is associated with the coexistence of phases of the types $La_2Co^{2+}Mn^{4+}O_6$ and $La_2Co^{3+}Mn^{3+}O_6$, whereas, in [6], the anomalous behavior in the dynamic susceptibility was interpreted within the model of separation into phases in which Co^{2+} and Mn^{4+} ions are in ordered and disordered states. Like classical manganites, these compounds can be doped with alkaline-earth elements. For example, in [9], the authors reported a synthesis of a $LaSrCo^{3+}Mn^{4+}O_6$ compound whose magnetic properties were analyzed within a model in which the moments of cobalt ions in an intermediate spin state are antiparallel to those of manganese ions.

In the present paper, we carry out X-ray, neutron-diffraction, and magnetic investigations of the base compound $La_2CoMnO_{6+\delta}$, as well as of compounds doped with strontium and calcium ions. The doping was performed in order to bring a part of Co^{2+} ions to a trivalent state. We show that the properties of these compounds should be described within an inhomogeneous model, based on the 2+ and 4+ valent states of cobalt and manganese ions, respectively.

2. EXPERIMENT

Samples of $La_2CoMnO_{6+\delta}$, $La_{1.4}Ca_{0.6}MnCoO_6$, and $La_{1.6}Sr_{0.4}CoMnO_6$ solid solutions were prepared by the conventional ceramic technology in air at T = 1320°C for a long period of time to improve the homogeneity of the chemical composition and slowly cooled to room temperature. By thermogravimetric analysis, it was established that the parent compound $La_2CoMnO_{6+\delta}$ had an excess of oxygen. Its chemical formula was

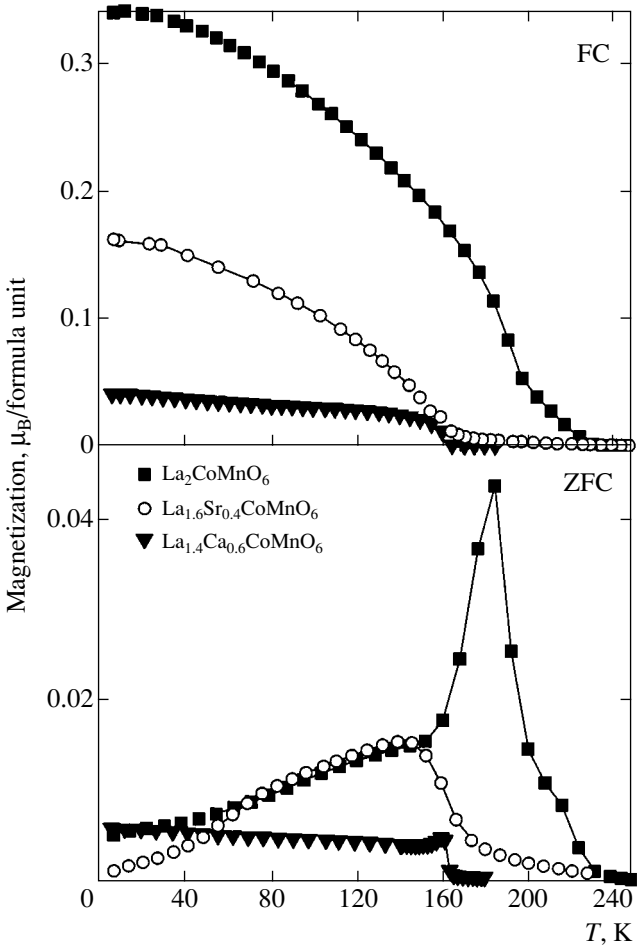


Fig. 1. ZFC- and FC-magnetizations of La_2CoMnO_6, $La_{1.4}Ca_{0.6}CoMnO_6$, and $La_{1.6}Sr_{0.4}CoMnO_6$ samples as a function of temperature in a field of H = 100 Oe.

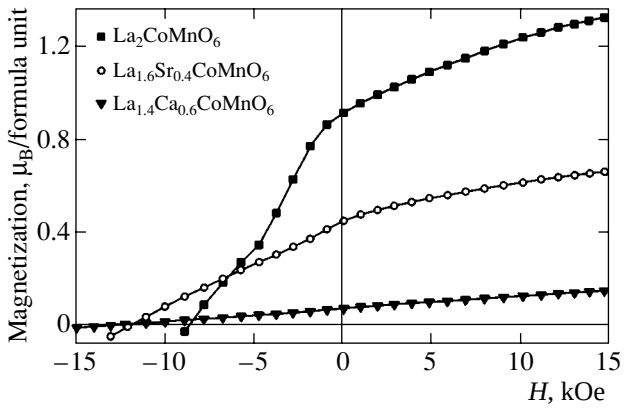


Fig. 2. Magnetization of La_2CoMnO_6, $La_{1.4}Ca_{0.6}CoMnO_6$, and $La_{1.6}Sr_{0.4}CoMnO_6$ samples as a function of an external magnetic field at a temperature of T = 7 K.

$La_2CoMnO_{6.1}$. Therefore, this compound was reduced in vacuum (1000°C, 24 hours) to a stoichiometric compound of La_2CoMnO_6. An X-ray phase analysis carried out on a DRON-3M diffractometer with the K_α radiation of Cu did not reveal any traces of extraneous phases. The electric conductivity was measured by the standard four-probe technique. The magnetization was measured on a commercially available QI-3001 vibration magnetometer. Neutron-diffraction investigations with a wavelength of λ = 1.7974 Å were carried out with a FIREPOD diffractometer at the Berlin Neutron Scattering Center (BENSC). The crystalline and magnetic structures were determined with the use of the FullProf program.

3. RESULTS AND DISCUSSION

Figure 1 represents the temperature dependence of the magnetization measured during heating in a magnetic field of 100 Oe after zero-field cooling (ZFC) and after field cooling (FC) in the same field. The figure shows that the FC magnetization is several times greater than the ZFC magnetization, especially at low temperatures. This feature is characteristic of cobaltites and lightly doped manganites and is attributed to the large magnetic anisotropy of these substances [10]. The second characteristic feature manifests itself in the anomalous behavior of magnetization below the Curie temperature. However, compounds doped with Ca and Sr ions show a rather sharp transition to a magnetically ordered state at lower temperatures (T_C ~ 160–170 K). The measurements of magnetization as a function of applied field have confirmed that these substances have large magnetic anisotropy. The saturation of magnetization at T = 7 K was not observed up to a field of 16 kOe, and the coercivity was as large as 10–15 kOe (Fig. 2). These measurements do not allow one to evaluate the spontaneous magnetization of samples; however, it should be noted that the substitution of strontium or calcium ions for lanthanum has led to a rather sharp decrease in magnetization.

The measurements of electric conductivity have shown that all compounds exhibit semiconductor-type resistivity versus temperature. No pronounced anomalies were observed near the magnetic ordering point.

The crystalline structure of a $La_2CoMnO_{6.1}$ sample was determined using powder X-ray patterns obtained at room temperature. Good values of the reliability factor were obtained in a model where a $La_2CoMnO_{6.1}$ sample was calculated as a mixture of two phases with perovskite structures. In this model, the content of the orthorhombic phase (*Pbnm* space group) amounts to 75%, whereas the rhombohedral phase ($R\bar{3}c$ space group) amounts to 25%.

The neutron-diffraction patterns of a La_2CoMnO_6 sample were obtained at room and liquid-helium temperatures. A stoichiometric vacuum-annealed sample contained only the orthorhombic phase. However, the best agreement between calculated and experimental data was again obtained in the two-phase model (Fig. 3). In this model, a sample was considered as a

mixture of two phases: one of the space group *Pbnm* and the other of the space group $P2_1/n$. In the space group *Pbnm*, cobalt and manganese ions occupy the same positions, whereas in the space group $P2_1/n$, they occupy different positions T(1) and T(2) in the crystalline structure.

The fraction of the phase in which cobalt and manganese ions are ordered amounted to 27% at helium temperature. The typical angles and lengths of bonds for the two phases are shown in the table. Note that the lengths of the T(2)–O bonds for the phase with ordered cobalt and manganese ions are about 1.89 Å on average, which is characteristic of Mn^{4+} ions, whereas the lengths of the T(1)–O bonds, whose mean value is 2.04 Å, are characteristic of cobalt ions in the bivalent state. This sample was also calculated within a homogeneous model (the space group $P2_1/n$) with regard to antisite defects. The calculation suggests that there are about 30% of such defects; i.e., 70% of T(1) positions are occupied by cobalt ions, and 30%, by manganese ions. The converse is true for the T(2) positions as well. If there were no antisite defects (the case of total ordering), then all Co ions would occupy the T(1) positions and Mn ions would occupy the T(2) positions. However, the reliability factors in this model are somewhat worse than those in the two-phase model. The mean value of the magnetic moment per formula unit is $\langle CoMn\rangle = 5\mu_B$, which is in good agreement with the expected value for cobalt and manganese ions in the 2+ and 4+ valent states, respectively.

The doping with strontium ions brought the whole sample to the rhombohedral phase. The crystalline structure was calculated in two models: in a single-phase model (the space group $R\bar{3}$) and a two-phase model (one phase is of the space group $R\bar{3}c$ and the other, of the group $R\bar{3}$). In the space group $R\bar{3}$, cobalt and manganese ions occupy different positions, whereas, in the space group $R\bar{3}c$, they occupy the same position. In the two-phase model, we could not calculate the coordinates of ions in the ordered phase because the content of this phase was no greater than 5%. In the model with antisite defects, the ratio of the occupation numbers of positions was 60 to 40%.

The calculated and measured neutron-diffraction patterns for $La_{1.6}Sr_{0.4}CoMnO_6$ are shown in Fig. 4. The lengths of bonds in the model with antisite defects are shown in the table. A decrease in the mean values of the lengths of bonds in octahedra and increased disorder in the arrangement of cobalt and manganese ions are attributed to the increased number of trivalent ions of cobalt due to the substitution of two-valent strontium ions for lanthanum ions.

In this sample, the contribution of magnetic scattering of neutrons proved to be much less than that in the case of a stoichiometric compound and the calculated magnetic moment per formula unit was about $1.2\mu_B$.

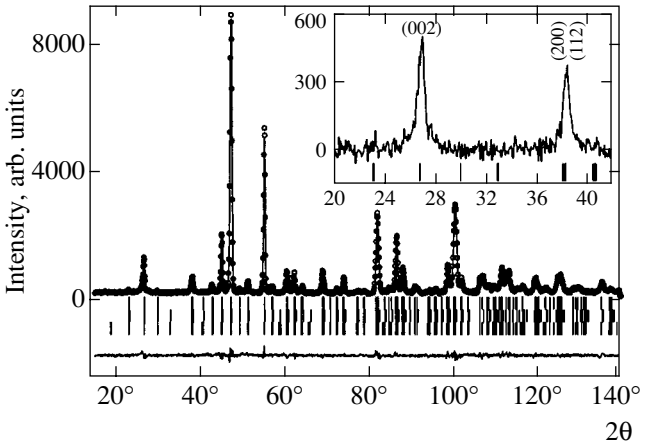


Fig. 3. Refinement of the La_2CoMnO_6 structure in the two-phase model by the Rietveld method using the neutron-diffraction patterns obtained at $T = 1.5$ K; (circles) experimental data, (solid line) computed curve, and (lower solid line) their difference. Vertical segments indicate the computed positions of 2θ reflexes; the upper row of segments corresponds to the orthorhombic crystalline phase, the middle row, to the monoclinic crystalline phase, and the lower row, to the magnetic phase. The inset shows the magnetic contribution to certain reflexes.

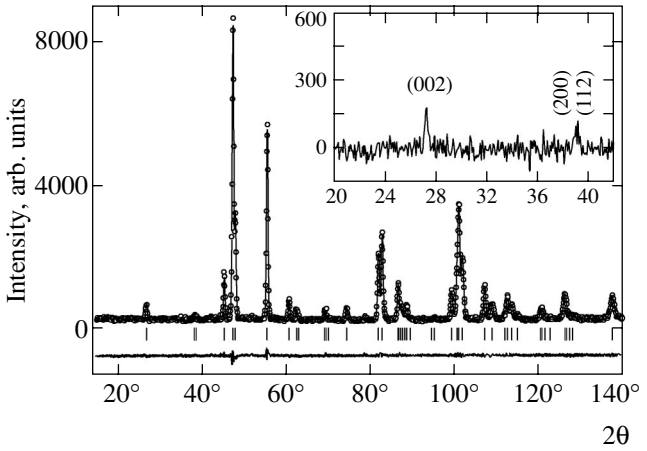


Fig. 4. Refinement of the $La_{1.6}Sr_{0.4}CoMnO_6$ structure in the two-phase model by the Rietveld method using the neutron-diffraction patterns obtained at $T = 298$ K; (circles) experimental data, (solid line) computed curve, and (lower solid line) their difference. Vertical segments indicate the computed positions of 2θ reflexes. The inset shows the magnetic contribution to certain reflexes.

Conversely, the substitution of calcium ions for lanthanum stabilized the orthorhombic structure. The crystalline structure of $La_{1.4}Ca_{0.6}CoMnO_6$ was calculated both in the two-phase and single-phase models (Fig. 5). Just as in the previous cases, the two-phase model gives somewhat better reliability factors. However, the features of the neutron-diffraction patterns associated with

The results of calculation of the crystalline structure and the bond lengths for various compounds. T(1) is a position predominantly occupied by cobalt ions and T(2) is a position predominantly occupied by manganese ions

Compound	La_2CoMnO_6		$La_{1.4}Ca_{0.6}CoMnO_6$	$La_{1.6}Sr_{0.4}CoMnO_6$
Temperature	1.5 K		298 K	298 K
Space group	*Pbnm*	$P2_1/n$	$P2_1/n$	$R\bar{3}$
a, Å	5.522921	5.513146	5.442330	5.475994
b, Å	5.484688	5.462361	5.410093	–
c, Å	7.777248	7.755850	7.656228	13.243181
β	–	90.025146°	90.029251°	–
La/A: *x*	0.99900	0.98951	0.99838	0
La/A: *y*	0.02197	0.02569	0.01791	0
La/A: *z*	1/4	0.23905	0.24955	0.25074
T(1): *x*	1/2	1/2	1/2	0
T(1): *y*	0	0	0	0
T(1): *z*	0	0	0	0
T(2): *x*	–	0	0	0
T(2): *y*	–	1/2	1/2	0
T(2): *z*	–	0	0	1/2
O(1): *x*	0.06911	0.05530	0.06036	0.54445
O(1): *y*	0.49333	0.49376	0.49490	0.00093
O(1): *z*	1/4	0.25440	0.25311	0.24785
O(2): *x*	0.72138	0.71270	0.72922	–
O(2): *y*	0.27579	0.27088	0.27007	–
O(2): *z*	0.03794	0.02209	0.03113	–
O(3): *x*	–	0.74421	0.72667	–
O(3): *y*	–	0.28573	0.27642	–
O(3): *z*	–	0.46603	0.47126	–
R_{Bragg}, %	2.98	4.19	5.02	5.45
χ^2, %	1.94	1.94	1.93	2.01
R_{Magn}, %	5.91	5.91	–	–
T(1)–O(1), Å	1.983 × 2	2.078 × 2	1.965 × 2	1.956 × 6
T(1)–O(2), Å	1.978 × 2	2.026 × 2	1.950 × 2	–
T(1)–O(3), Å	1.969 × 2	1.997 × 2	1.943 × 2	–
⟨T(1)–O⟩, Å	1.977	2.034	1.952	1.956
T(2)–O(1), Å	–	1.929 × 2	1.935 × 2	1.930 × 6
T(2)–O(2), Å	–	1.896 × 2	1.929 × 2	–
T(2)–O(3), Å	–	1.851 × 2	1.918 × 2	–
⟨T(2)–O⟩, Å	–	1.892	1.927	1.930

the ordering of Co^{2+} and Mn^{4+} ions are still less pronounced than in the sample doped with strontium ions. Therefore, the results of calculations are presented for the model with antisite defects (see table). In the calcium-containing sample, we could not distinguish a coherent magnetic scattering of neutrons; however, magnetic measurements show that magnetic ordering in this sample occurs in a rather narrow range of tem-

peratures near $T = 160$ K, which is not characteristic of spin glasses but is quite possible in a two-phase magnetic state when the content of ferromagnetic phase is very small.

As we pointed out in the Introduction, valent states of cobalt and manganese ions in La_2CoMnO_6 are described in different models [6, 7]. According to [7], a low-temperature ferromagnetic phase that is ordered at $T = 160$ K corresponds to the ordering of Co^{2+} and Mn^{4+} magnetic ions, while a high-temperature phase that is ordered at $T = 230$ K, corresponds to the distribution of valences of Co^{3+} and Mn^{3+} ions. In this phase, cobalt ions must occupy a low-spin state because, according to magnetic measurements, the spontaneous magnetization per formula unit is about $4\mu_B$. According to [7], La_2CoMnO_6 samples consist of a mixture of these two phases. Objections against this model arise from the fact that cobalt ions in the low-spin state are diamagnetic ($S = 0$) and therefore cannot take part in exchange interactions. For example, Pb_2MnNbO_6-type perovskites, in which manganese is trivalent and niobium ions are pentavalent and are in the diamagnetic state, represent spin glasses with low freezing temperature (T_f = 40 K) of the spin moments of manganese ions [11]. However, according to our neutron-diffraction studies, both phases—with ordered and disordered cobalt and manganese ions—are ferromagnetic. Our neutron-diffraction patterns are in agreement with the results of electron-microscope studies of fine-dispersed La_2CoMnO_6 [5], where the authors showed that only 5% of grains exhibit superstructural reflexes associated with the ordering of cobalt and manganese ions. They also made an assumption that the ordered phase has a lower Curie temperature compared with the disordered phase; however, this assumption is incorrect from the viewpoint of the mechanism of exchange interaction between cobalt and manganese ions. It is well known that the Co^{2+}–O–Co^{2+} and Mn^{4+}–O–Mn^{4+} exchange interactions are antiferromagnetic [12]; therefore, the antisite defects in the ordered phase of La_2CoMnO_6 should reduce the Curie temperature or should even give rise to an antiferromagnetic component. Therefore, we assume that low-temperature anomalies in the magnetic properties of La_2CoMnO_6-type compounds are associated with a crystallographic phase that is disordered with respect to 3*d* ions.

A question arises as to what spin state the trivalent cobalt ions are in that arise under the doping of La_2CoMnO_6 with alkaline-earth ions or under the excess of cobalt ions over manganese ions. Recall that, in $LnCoO_3$ (Ln is a lanthanide) perovskites, Co^{3+} ions are in the low-spin diamagnetic state at low temperatures and gradually go over to the intermediate-spin state as temperature increases [13]. Judging by the magnetic properties, the doping of $LaCoO_3$ with manganese ions gives rise to Co^{2+}–Mn^{4+} pairs in the matrix of Co^{3+} ions that are in the low-spin state [14]. The

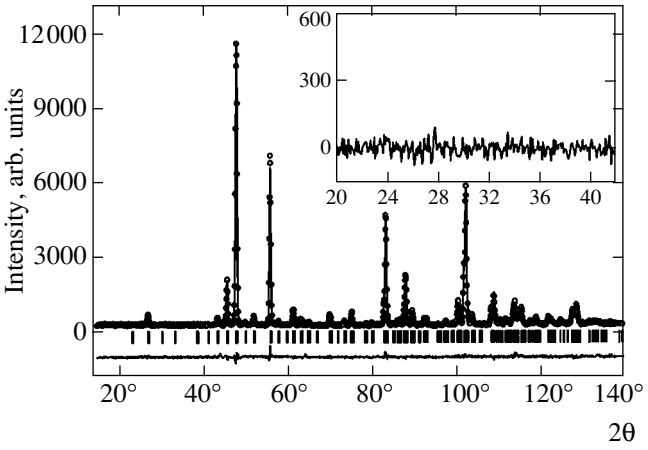


Fig. 5. Refinement of the $La_{1.4}Ca_{0.6}CoMnO_6$ structure in the two-phase model by the Rietveld method using the neutron-diffraction patterns obtained at $T = 298$ K; (circles) experimental data, (solid line) computed curve, and (lower solid line) their difference. Vertical segments indicate the computed positions of 2θ reflexes. The magnetic contribution could not be distinguished (see the inset).

magnetic properties of La_2CoMnO_6 doped with alkaline-earth ions can also be consistently explained within the model of an inhomogeneous magnet in which Co^{3+} ions are in the low-spin state. In our opinion, these solid solutions consist of clusters with different concentrations of strontium: the ground state corresponds to microscopic domains with the chemical composition close to that of La_2CoMnO_6, whereas clusters with increased content of strontium correspond to the spin-glass-type magnetic ground state. The clusters must be sufficiently large since the transition to the paramagnetic state may be sufficiently sharp and it occurs at a sufficiently high temperature. The magnetic behavior of La_2CoMnO_6 can also be accounted for based on the assumption that a sample consists of microscopic domains with different concentrations of cobalt (manganese) ions.

The model explaining magnetic phase separation in $La_{1-x}Sr_xMnO_3$ manganites and $La_{1-x}Sr_xCoO_3$ cobaltites via the intrinsic chemical inhomogeneity of solid solutions is corroborated by X-ray absorption fine structure (XAFS) measurements [15], by high-resolution electron-microscope data [16], as well as by other spectroscopic investigations [17]. For instance, according to [15], $La_{1-x}Sr_xMnO_3$-type lightly doped manganites consist of strongly distorted domains with a structure close to that of $LaMnO_3$ and of weakly distorted domains that are doped with strontium ions. A similar picture is observed in cobaltites. According to NMR studies, the phase corresponding to the low-spin state of Co^{3+} ions was observed even under a 50%-substitution of strontium ions for lanthanum ions [4]. Recall that a transition to the ferromagnetic state in cobaltites

occurs when the concentration of strontium ions approaches 20%.

Investigations carried out with single crystals or with samples prepared with the use of special chemical techniques in which precursor atoms are mixed on the nanometer scale do not significantly change the results of [15]. Therefore, we are dealing with the intrinsic structural and chemical inhomogeneity of solid solutions, which, apparently, cannot not be removed in principle.

Most likely, $LaCo_{1-x}Mn_xO_3$ solid solutions are not homogeneous objects either, but consist of two, or even three, different magnetic phases. We assume that the ordering of cobalt and manganese ions occurs only in a rather narrow range of concentrations near $x = 0.5$. In this case, the main part of a sample may consist of a phase with disordered cobalt and manganese ions, which contains clusters with a deficiency or an excess of cobalt with respect to the value of $x = 0.5$. The disordered phase with deficiency of cobalt is a ferromagnetic phase with a characteristic Curie temperature of 160–170 K. A significant contribution to the ferromagnetism of this phase is made by the positive exchange interaction between Mn^{3+} and Mn^{4+} ions. The magnetoresistance effect observed in $LaCo_{1-x}Mn_xO_3$ [6] is likely to be primarily associated with this phase. The phase with ordered arrangement of Co^{2+} and Mn^{4+} is characterized by the highest T_C, since the geometry of the Co^{2+}–Mn^{4+} exchange bonds is optimal: according to the Blasse model [18], each Co^{2+} ion positively interacts with six Mn^{4+} ions and vice versa. However, the phase with an increased concentration of cobalt ions is not ferromagnetic; an interaction between this phase and the phases with a deficiency in cobalt leads to frustration of exchange bonds on the interface between phases, which makes a significant contribution to magnetic anisotropy. The coexistence of orthorhombic and rhombohedral phases in nonstoichiometric $La_2CoMnO_{6+\delta}$ is most likely associated with the inhomogeneous distribution of oxygen over a sample. The domains with increased oxygen content are rhombohedral, whereas the domains with decreased oxygen content are characterized by orthorhombic distortions of a unit cell. This situation is largely similar to that in stoichiometric lanthanum manganite. In $LaMnO_{3+\delta}$, oxygen nonstoichiometry is due to the vacancies of lanthanum and manganese cations; an increased oxygen content leads to a transition from the orthorhombic to the rhombohedral structure [19]. Like in the case of lanthanum manganite, the substitution of strontium for lanthanum in La_2CoMnO_6 stabilizes the rhombohedral structure, whereas the substitution of calcium for lanthanum stabilizes the orthorhombic structure. However, the mechanisms of exchange interaction in $LaMnO_{3+\delta}$ and La_2CoMnO_6 are completely different. In $LaMnO_{3+\delta}$, ferromagnetism is associated with the removal of static orbital distortions due to doping with Mn^{4+} ions [20], whereas La_2CoMnO_6 is not a Jahn–Teller magnet; its ferromagnetic structure is attributed to the positive exchange interaction between Co^{2+} and Mn^{4+} ions, and doping with alkaline-earth metals leads to a gradual destruction of ferromagnetism.

4. CONCLUSIONS

The X-ray, neutron-diffraction, and magnetic measurements carried out on $La_2CoMnO_{6+\delta}$ and $La_{2-x}A_xCoMnO_6$ (A = Ca and Sr) systems have revealed specific features of the magnetic state of these systems and its relation to the crystalline structure. It has been shown that the base compound La_2CoMnO_6 remains homogeneous in a rather wide range of oxygen concentrations; the oxidation leads to a transition from the orthorhombic to the rhombohedral structure. The rhombohedral structure stabilizes under the doping with strontium, whereas the orthorhombic structure stabilizes under the doping with calcium as in lanthanum manganite. The type of distortions of the crystalline structure is attributed to the size effect, i.e., associated with the variance of cation sizes at different positions. The analysis of the crystalline structure shows that the cobalt and manganese ions tend to ordering with the ratio 1 : 1, and all samples consist of ordered and disordered regions to a certain extent. The most probable reason for the phase separation lies in local fluctuations of the chemical composition.

According to analysis of bond lengths, cobalt and manganese ions in the ordered phase have valences of 2+ and 4+, respectively. Doping with alkaline-earth ions leads to a transition of cobalt ions from a bivalent to a trivalent state, which is most likely a low-spin state ($S = 0$). In this case, compounds are separated into Co^{3+}-rich and Co^{3+}-poor regions in which spin-glass-type and ferromagnetic states are realized, respectively.

ACKNOWLEDGMENTS

This work was supported by the Foundation for Basic Research of the Republic of Belarus, project no. F03MS-021 and under the State Program of Oriented Basic Research “Nanomaterials and Nanotechnologies,” Task 3.3.

REFERENCES

1. K. Kumagai, A. Iwai, Y. Tomioka, *et al.*, Phys. Rev. B **59**, 97 (1999).
2. V. Podzorov, B. G. Kim, V. Kiryukhin, *et al.*, Phys. Rev. B **64**, 140406(R) (2001).
3. E. Dagotto, T. Hotta, and A. Moreo, Phys. Rep. **344**, 1 (2001).
4. P. L. Kuhns, M. J. R. Hoch, W. G. Moulton, *et al.*, Phys. Rev. Lett. **91**, 127202 (2003).
5. R. Mahendiran, Y. Breard, M. Hervieu, *et al.*, Phys. Rev. B **68**, 104402 (2003).

6. I. O. Troyanchuk, L. S. Lobanovsky, D. D. Khalyavin, *et al.*, J. Magn. Magn. Mater. **210**, 63 (2000).
7. P. A. Joy, Y. B. Khollam, and S. K. Date, Phys. Rev. B **62**, 8608 (2000).
8. V. L. J. Joly, P. A. Joy, S. K. Date, and C. S. Gopinath, J. Phys.: Condens. Matter **13**, 649 (2001).
9. J. Androulakis, N. Katsarakis, J. Giapintzakis, *et al.*, J. Solid State Chem. **173**, 350 (2003).
10. R. Ganguly, A. Maignan, C. Martin, *et al.*, J. Phys.: Condens. Matter **14**, 8595 (2002).
11. E. E. Havinga, Philips Res. Rep. **21**, 432 (1966).
12. A. L. Cornelius and B. E. Light, Phys. Rev. B **68**, 014403 (2003).
13. C. Maris, Y. Ren, V. Volotchaev, *et al.*, Phys. Rev. B **67**, 224423 (2003).
14. R. I. Dass and J. B. Goodenough, Phys. Rev. B **67**, 014401 (2003).
15. T. Shibata, B. Bunker, J. F. Mitchell, and P. Schiffer, Phys. Rev. Lett. **88**, 207205 (2002).
16. G. Baio, G. Barucca, R. Caciuffo, *et al.*, J. Phys.: Condens. Matter **12**, 9761 (2000).
17. J. C. Loudon, N. D. Mathur, and P. A. Midgley, Nature **420**, 797 (2002).
18. G. Blasse, J. Phys. Chem. Solids **26**, 1969 (1965).
19. I. O. Troyanchuk, V. A. Khomchenko, A. N. Chobot, and H. Szymczak, J. Phys.: Condens. Matter **15**, 6005 (2003).
20. I. O. Troyanchuk, O. S. Mantytskaya, A. N. Chobot, and H. Szymczak, Zh. Éksp. Teor. Fiz. **122**, 347 (2002) [JETP **95**, 300 (2002)].

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 99, No. 2, 2004, pp. 370–378.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 2, 2004, pp. 415–425.

SOLIDS
Electronic Properties

Observation of Latent Coherent Effects in Randomized Systems

A. I. Golovashkin, A. N. Zherikhin, L. N. Zherikhina, G. V. Kuleshova, and A. M. Tskhovrebov

Lebedev Physical Institute, Russian Academy of Sciences, Moscow, 119991 Russia
e-mail: tshovrebov@rambler.ru

Received February 2, 2004

Abstract—Coherent interference effects of the following three types are experimentally discovered in disordered (randomized) systems: (i) Josephson behavior of the HTSC polycrystal BaKBiO in the phase-separated state; (ii) oscillations of bismuth film resistance, which are periodic in "direct" magnetic field; and (iii) mesoscopic oscillations of the resistance in the course of film growth. In the first case, the method for detecting the "latent" nonstationary Josephson effect is substantiated by the frequency modulation method for microwave radiation, while in the other two cases, simple models are proposed to explain the nature of coherent oscillations of the resistance. The analogy between the observed oscillations and the Josephson effect in randomized systems is discussed. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

One of the most convincing proofs of the Josephson nature of an object under investigation is the observation of the nonstationary Josephson effect on this object since the nonstationary effect is a direct consequence of the coherent interaction of weakly coupled superconductors. To substantiate this, use is normally made of external manifestations of the Josephson effect, e.g., the emergence of singularities on the current–voltage (IV) characteristics of the sample or the suppression of the critical current under the action of microwave radiation or a weak magnetic field. However, as a rule all this can be treated as an indirect and insufficient confirmation of the Josephson nature of the object. Indeed, the critical current attributed to a weak Josephson coupling might be just the small macroscopic critical current of the sample. The variation of the IV characteristic under the action of radiation is infrequency due to microwave detection of nonohmicity in the sample junctions. The suppression of the critical current by a weak magnetic field also turns out to be a result of the low value of the critical field of the superconductor; similar step singularities in principle can appear on the IV characteristic of the sample during vortex detachment induced by microwave radiation. In contrast to the external manifestation described above, the nonstationary effect is a direct consequence of coherent properties of weakly coupled superconductors; the observation of this effect provides a reliable proof of the Josephson nature of the system in question. However, the detection of the nonstationary Josephson effect from the presence of Shapiro steps on the IV characteristics is possible either for an individual junction or for a system of identical tunnel junctions. A considerable spread in the values of the critical current, normal resistance, and other parameters in a system of multiple Josephson junctions smears the Shapiro steps and rules out observation of the nonstationary effect by a conventional method in a sample containing such junctions. For this reason, a strong broadening of the spectrum of self-radiation generated by a junction in a system of this kind complicates the detection of such a noiselike signal by a narrowband receiver. At the same time, knowledge of the nature of weak bonds operating in randomized systems or in systems with a considerable spread in parameters of such bonds is essential in a number of cases.

In the present work, we propose a method for detecting the "latent" nonstationary Josephson effect in randomized systems and consider the experimental results on observation of coherent interference effects in the disordered superconducting system BaKBiO and ultrathin bismuth films with random violations of uniformity of the parameters.

2. FREQUENCY MODULATION METHOD

In spite of a considerable spread in weak coupling parameters, a manifestation of the nonstationary Josephson effect in randomized superconducting systems can be observed by synchronizing the change in the IV characteristics of an individual junction by external action. If we periodically change the frequency of the irradiating microwave field, the potential difference across each junction will vary synchronously with the variation of the frequency of incident radiation (naturally, if the junctions indeed exhibit the Josephson properties and the current passed through these junc-

tion is strong enough, $I > I_c^{(J)}$, where $I_c^{(J)}$ is the Josephson critical current). Thus, the Fourier component of the voltage drop across the sample, detected by a narrowband ac voltmeter or a synchronous detector and having a frequency corresponding to the period of frequency oscillations in microwave radiation, directly indicates the presence of latent Josephson properties in the given system.

In a practical realization of the proposed method, one encounters a spurious amplitude modulation of the supplied microwave radiation. Amplitude modulation inevitably appears when a frequency-modulated microwave signal passes through a waveguide tract with nonideal amplitude-frequency characteristics (AFCs) when the signal is incident on a sample located at the antinode of the electric field, which corresponds to a quite definite wavelength, or simply due to nonuniformity of the AFCs of a real microwave oscillator. The potentialities of the output power automated control system (which is used, for example, in AFC panoramic displays) are limited by the amplitude-frequency properties of the tracking detector. For this reason, we take the contribution of the spurious amplitude modulation to the voltage drop detected on the sample and compare it with the contribution from the frequency modulation.

Let us qualitatively consider the response of a multiple system of nonidentical junctions on the frequency and amplitude modulation of incident microwave radiation. Obviously, a multiple Josephson system can be visualized as a 3D percolation network of tunnel junctions with different critical currents. It is well known that irradiation of an individual Josephson junction leads to the formation of an equidistant (in voltage) vertical segment (Shapiro steps) on the IV characteristic of the junction.

In the simplest model, the Shapiro steps must possess an infinitely large differential conductivity; in other words, Shapiro steps are a sort of superconducting segments of the IV characteristic [3]. In a 3D body, in accordance with Ohm's law, the transport current passes along the trajectories corresponding to maximal conductances. Thus, the current "tries to pass" via the superconducting regions; i.e., it is concentrated by "superconducting crosspieces." An analogous requirement determining the choice of geometrical paths for the current in the case of an infinitely large differential conductivity (vertical segments of the IV characteristic) can now be formulated for current increments. Any increment smaller than the current interval corresponding to a vertical segment of the IV curve is "short-circuited" by a junction of the circuit with an infinitely large differential conductivity. As soon as this current interval is exceeded, the system will try to find new "superconducting" junctions. Consequently, with a monotonic increase in the current, a randomized system strives to pass any current increment via subcircuits with an infinitely large differential conductivity, which corresponds to Shapiro steps in the case when the system is exposed to microwave radiation. The increment in the current density through the regions in which the working point of the IV characteristic lies on the Shapiro steps will be maximal, which is analogous to superconducting short-circuiting.

From the standpoint of the system's stability to random deviations, "superconducting regions for $V \neq 0$" are the most attractive (V is the potential difference across a Josephson junction). If we characterize each state of the system by a set of currents and voltages at all junctions of the circuit, the statistical weight of the states will be maximal when the working point of individual elements of the circuit lie on the segment of the IV characteristic with an infinitely large conductance. This result is obtained since one value of voltage on a "superconducting segment" of the IV characteristic with $V \neq 0$ corresponds to a set of current values (the segment of the IV curve is degenerate in current). Consequently, under the action of microwave radiation on a percolation Josephson network, the transport current chooses the paths crossing the Josephson junctions whose working points under the given conditions lie on the vertical segment of a Shapiro step. It is as if the multiple Josephson system as a whole tries to hold out on a step in each of its junctions. If the frequency of microwave radiation is varied, the voltage drop across all the Josephson junctions for which the working point is on the vertical segment of the IV curve (and, accordingly, across parallel-connected subcircuits and, hence, across potential contacts of a macroscopic sample) changes synchronously. In addition, the change in the voltage is a linear function of the change in frequency.

A change in the amplitude of microwave radiation results in a change in the height of the Shapiro step, the corresponding voltage remaining unchanged. Such an effect on the multiple Josephson system causes either a change in the position of the current through a Josephson junction with the voltage of the step unchanged, or a slippage of the working point from the step, and the system on the whole will try to find an optimal path for the current. In the latter case, the probability of slipping from the lower and upper edges of the step are compensated and the voltage drop on the average remains unchanged. In this case, the voltage might change only due to a change in the height of the zeroth step, which is quadratic in the amplitude of microwave radiation. In both cases, a considerable part of the junctions correspond to vertical segments of the IV curve and the total voltage across the potential contacts of the sample remains unchanged.

The pattern of the "frequency" and "amplitude" effects can be visualized as follows: a frequency variation causes a proportional change in the voltage across most of the junctions; i.e., the response of the system is considerable and linear. An amplitude variation does not lead to a significant variation in the voltage distribution in the system and causes a weak response quadratic in the amplitude variation (mainly near the zeroth

step). When a Josephson junction is exposed to microwave radiation, a change in the frequency affects the voltage drop across the junction more strongly than an amplitude variation (at least, as long as the voltage monotonically depends on these effects). Indeed, otherwise, an infinitely small alternating electromagnetic field of a very low frequency (e.g., a noise signal of any origin) could completely suppress the Josephson critical current, which would practically rule out the very possibility of observing the Josephson effect. The height $I_0^{(\text{step})}$ of the Josephson step is

$$I_0^{(\text{step})} = I_c^{(J)} J_0\left(\frac{2eU_0}{\hbar\omega}\right) \approx I_c^{(J)}\left(1 - \frac{2e^2U_0^2}{\hbar^2\omega^2}\right),$$

where $I_c^{(J)}$ is the Josephson critical current, J_0 is the zero-order Bessel function, U_0 and $\omega = 2\pi f$ are the amplitude and frequency of the microwave radiation, $2e$ is the charge of the Cooper pair, and $\hbar$ is the Planck constant.

Thus, the strong response of a multiple Josephson system to the frequency modulation of incident radiation (as compared to the response to the amplitude modulation) is a direct consequence of the Josephson properties of the junctions forming the system. This characteristic property can be used as direct evidence for the Josephson nature of the object under investigation. To detect such properties in actual practice, one should slowly scan the current through the sample to record, using a selective narrow-band voltmeter or a synchronous detector, the component of the voltage drop across the sample with a frequency corresponding to the period of frequency oscillations in the modulated microwave field acting on the sample. It is not necessary to take special measures to suppress spurious amplitude modulation since the above qualitative analysis shows that, under the natural condition

$$\frac{\hbar\omega}{\Delta} > \left(\frac{2eU_0}{\hbar\omega}\right)^2,$$

the response of a multiple Josephson system to amplitude modulation is smaller than the response to the frequency modulation of the microwave signal (Δ is the energy gap in the superconductor). Instead of suppressing the spurious amplitude modulation, we must compare the response of the system to frequency modulation, obtained as a function of the current passing through the sample, with an analogous function under the action of amplitude modulation of microwave radiation of a stable frequency. In both measurements, the modulation depth must be the same (otherwise, the signals should be reduced to the same level). A considerable excess of the former response over the latter in a wide range of currents directly indicates the Josephson nature of the multiple system studied even if the Shapiro steps on the static IV characteristic of the sample are unstable or blurred in view of the difference in the junctions.

3. EXPERIMENTAL RESULTS: HTSC BaKBiO

We used the above method for proving the Josephson nature of the system under investigation in experiments with polycrystals of the high-T_c superconductor $Ba_{0.6}K_{0.4}BiO_3$ (with a superconducting transition temperature of $T_c = 30$ K). Earlier, we discovered a number of anomalies in the behavior of this material at $T < T_c$ [4].

(i) A return of the resistive state is detected at temperatures below T_c, indicating the nonmonotonic temperature dependence of the critical current.

(ii) A nonmonotonic dependence of the resistance on the magnetic field was observed; the sample resistance increases by several orders of magnitude as compared to its value prior to the superconducting transition (R_N), after which the resistance decreases to R_N upon a further increase in the field.

(iii) Low-frequency inductive measurements of the susceptibility of the sample in a state with a finite conductivity down to complete destruction of this state indicate the presence of the superconducting phase in the bulk [5].

These results indicate the possibility of self-consistent coexistence of the superconducting and insulator phases in the given system. The presence of the insulator phase explains the increase in the resistance under the action of the magnetic field at temperatures below T_c. On the other hand, complete suppression of superconductivity by a magnetic field is accompanied by the destruction of the insulator phase also, which reduces the resistance. Gorbatsevich *et al.* [6] theoretically substantiated the above model, assuming that the current transport via insulator regions is carried out via Josephson tunneling, which necessitates searching for the Josephson effect in $Ba_{1-x}K_xBiO_3$.

To prove the existence of the effect in this system, we used frequency modulation of the microwave oscillator X1-24 with a carcinotron operating in the frequency range $f = 8$–10 GHz with automatic stabilization of the output voltage (Fig. 1). Response V_F of the sample was detected by a selective voltmeter (Unipan *n*-233) at a modulation frequency of $F = 17$ Hz as a function of the current passed through the sample. The frequency deviation δf amounted to 600 MHz and $\delta f/f = \delta\omega/\omega = 6\%$. Since the sample was in an unstable state and the noise level was rather high, we measured the difference δV_F between signal $V_F^{(\text{FM})}$ in the presence and V_F in the absence of frequency modulation of microwave radiation: $\delta V_F = V_F^{(\text{FM})} - V_F$. Signals $V_F^{(\text{FM})}$ and V_F in a digital code were fed via an interface directly to the temporary memory unit of a microcom-

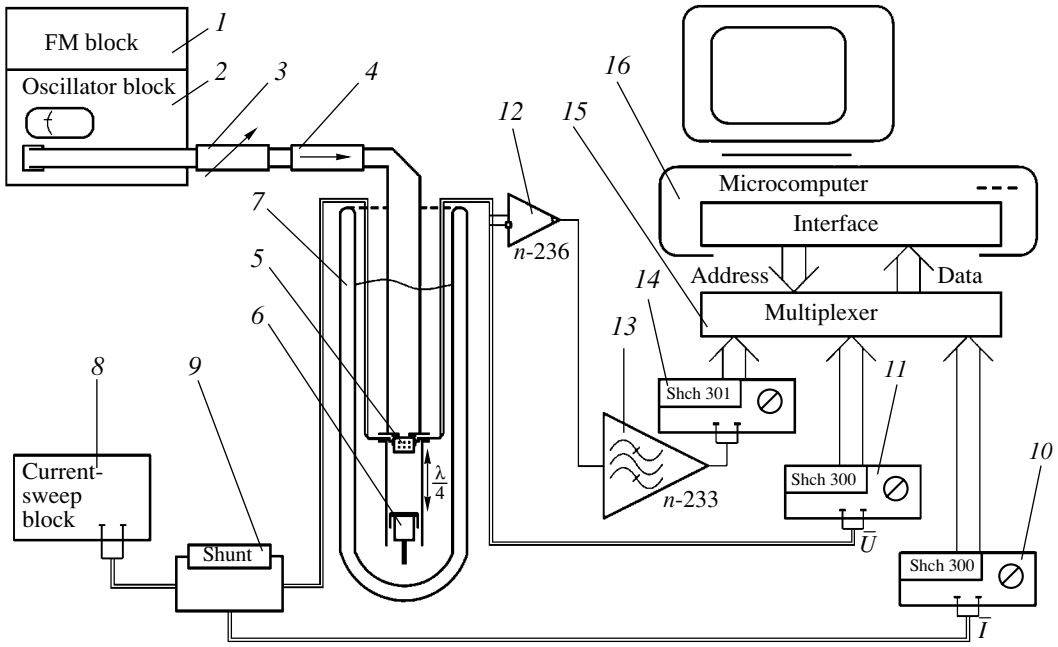


Fig. 1. Schematic diagram of the experiment with frequency modulation of the microwave oscillator: *1*—frequency modulation block of the microwave oscillator; *2*—oscillator block with a 3-cm backward wave tube; *3*—attenuator; *4*—microwave isolator, *5*—diaphragm with the sample at the joint between the waveguide and the tunable quarter-wave element; *6*—tuning piston; *7*—cryostat with liquid ^{4}He; *8*—slow-linear-sweep current generator; *9*—shunt for detecting the current introduced to the $Ba_{1-x}K_xBiO_3$ sample; *10*—voltmeter for numbering the running values of the current in the sample; *11*—voltmeter measuring the voltage drop for recording the IV characteristic of the sample; *12*—differential-input preamplifier for recording the ac voltage component across the sample; *13*—selective nanovoltmeter for separating the voltage component at the modulation frequency of the microwave radiation frequency; *14*—voltmeter for numbering the signal with the selective nanovoltmeter output; *15*—multiplexer for data input from three digital voltmeters to an ordinary 16-bit microcomputer bus; *16*—computer with a parallel interface.

puter, where the signal was accumulated in memory cells whose addresses were determined by the running current value through the sample and then the difference between the contents of the corresponding cells was calculated (symbols $\langle\ \rangle$ will be used to denote averaging of a signal accumulated in several paths). To make sure that the observed signal appears due to frequency modulation and not as a result of spurious amplitude modulation (caused by nonuniformity of the AFCs of the oscillator, waveguide, etc.), we compared $\langle \delta V_F^{(\mathrm{FM})} \rangle$ with an analogous signal $V^{(\mathrm{AM})}$ corresponding to specially introduced 100% amplitude modulation of the microwave radiation. Signal $\langle \delta V_F^{(\mathrm{AM})} \rangle$ was normalized to 6%, which corresponds to the majorant estimate of the depth of the spurious amplitude modulation in experiments with frequency modulation, $\langle \delta V_F^{(\mathrm{AM})} \rangle = 0.06 \langle \delta V_F^{(\mathrm{AM})} - V_F \rangle$. The value of $\langle \delta V_F^{(\mathrm{AM})} \rangle$ obtained in this way was then compared with $\langle \delta V_F^{(\mathrm{FM})} \rangle$. A comparison of $\delta V_F^{(\mathrm{AM})}$ and $\delta V_F^{(\mathrm{FM})}$ makes it possible to demonstrate that the signal $V_F^{(\mathrm{FM})}$ being measured is associated with frequency modulation alone and directly indicates the Josephson nature of the $Ba_{1-x}K_xBiO_3$ system in the resistive state at $T < T_c$.

Figure 2 shows the results of measurements of the response of a $Ba_{1-x}K_xBiO_3$ sample to modulation of microwave radiation and the IV characteristics of the sample recorded simultaneously with the measurements. It can be seen from the curves that a nonzero response of the sample was detected in the current interval $I_c < I < I_{c1}$, where I_{c1} [5] is the current corresponding to the beginning of the linear segment of the IV curve. It can be noted that the peak of the response of the given system to amplitude modulation of radiation lies in the range of small currents; this is in accordance with the statement that the value of $V_F^{(\mathrm{AM})}$ should be estimated taking into account only the contribution from the zeroth step, assuming that the contributions of the remaining steps are mutually compensated. A considerable excess in the sample response to frequency modulation (Fig. 2b) over the response to amplitude modulation (Fig. 2c) in a wide range of currents ($I_c < I < I_{c1}$) directly indicates the Josephson nature of the multiple system studied here in spite of the absence of stable Shapiro steps on the static IV characteristic of the sample. The asymmetry in the dependences of the

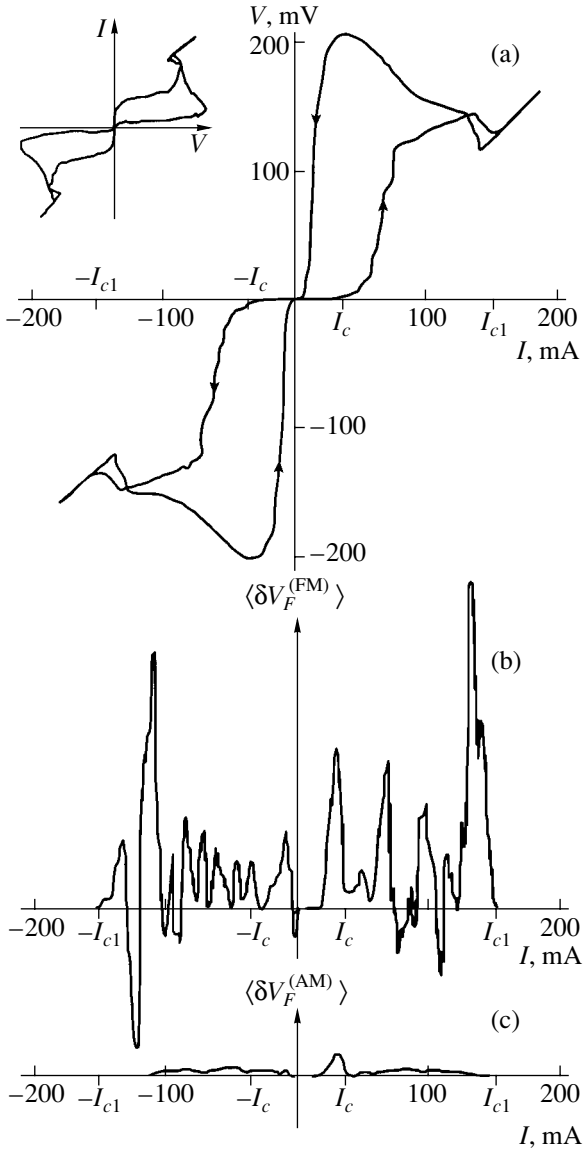


Fig. 2. (a) IV characteristic of the $Ba_{1-x}K_xBiO_3$ sample; the inset shows the IV characteristic of a sample with a traditional arrangement of the axes; (b) averaged component $\langle \delta V_F^{(FM)} \rangle$ of the sample response to periodic variation of the microwave radiation frequency, detected at modulation frequency F; (c) averaged response $\langle \delta V_F^{(AM)} \rangle$ to amplitude modulation with a depth reduced to the level at which frequency modulation deviates (6%).

responses $\langle \delta V_F^{(FM)} \rangle$ and $\langle \delta V_F^{(AM)} \rangle$ on the current passed through the sample is due to the fact that the microwave response was averaged separately for two directions of motion of the working point along the hysteresis IV curve of the sample. Figure 2 shows responses corresponding to increasing current passed through the sample.

The IV characteristic recorded under microwave irradiation at a stable frequency (without frequency and amplitude modulations) demonstrates the absence of conventional phase synchronization in the entire sample. Such a synchronization of Josephson junctions would lead in this case to the emergence of "giant" steps on the IV curve, which was not observed in experiment. The absence of such steps also rules out the effects like those of synchronous motion of vortices in the sample under the action of the microwave radiation.

4. AHARONOV–BOHM EFFECT AND RESISTANCE OSCILLATIONS IN MESOSCOPIC OBJECTS

The observation of quantum interference in a magnetic field (i.e., the Aharonov–Bohm effect) provides another opportunity for a direct substantiation of coherent properties of a randomized system. Unfortunately, such experiments in superconducting systems with a weak coupling can be carried out only in special cases: two tunnel junctions in a superconducting ring, an extended homogeneous Josephson junction, a network of identical DC-SQUIDs, etc.) In randomized Josephson systems, interference in a magnetic field cannot be observed since an appropriate synchronizing mechanism is not available. However, there exists a mechanism of self-consistent enhancement of the contribution to the overall effect from individual elements in a randomized ensemble of mesoscopic objects. Owing to this mechanism, the contribution becomes macroscopically observable. Here, we apply the term mesoscopic objects to objects of macroscopic size, which are in a nonsuperconducting state at a low temperature, when the mean free path of normal charge carriers and the phase coherence length are commensurate to the size of the object. It is well known [7, 8] that mesoscopic objects can exhibit interference effects.

The mesoscopic objects in which quantum interference is observed are microscopic structures based on high-quality metal films (characterized by a large phase coherence length) with a special geometry on the submicrometer scale (e.g., a ring or a cylinder with a diameter smaller than the coherence length of charge carriers). The area bounded by the ring or the cylinder base determines the period of resistance oscillations in a magnetic field ($\Delta B = \Phi_0^{(ne)}/S = 2\pi\hbar/neS$, where $\Phi_0^{(ne)}$ is the magnetic flux quantum in a system with n-electron correlations). However, we came across oscillations of the Hall resistance $R_{xy}(B)$ in Bi films [9] intended for the observation of the quantum Hall effect, i.e., with a geometry far from that of a solitary ring of submicrometer size. Obviously, in our experiments, a ring (or a microscopic region possessing similar interference properties) appeared in the film spontaneously (probably, during laser sputtering) [10]. The conductivity in such a microscopic region is high since interference effects require a large coherence length and, hence, the mean free path must be sufficiently large; this, in turn, leads to a high conductivity. The high local conductiv-

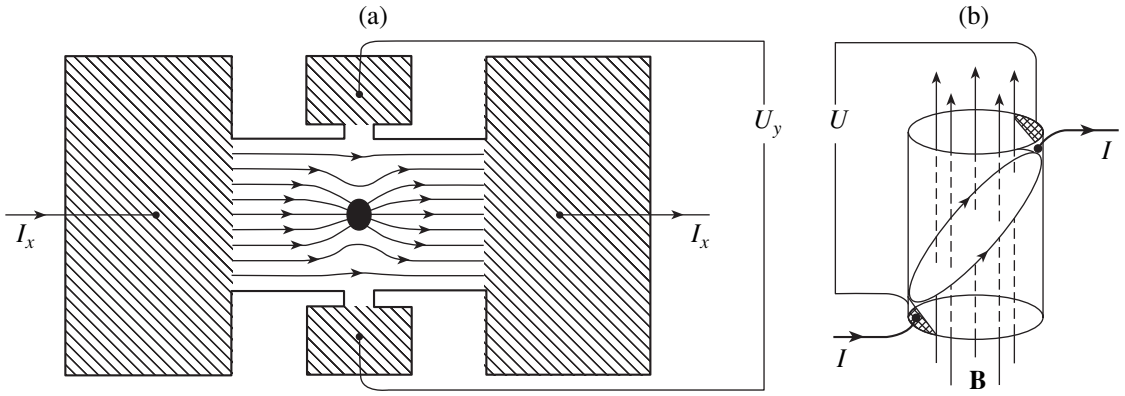


Fig. 3. (a) Schematic diagram of concentration of current lines by the high conductivity of a microregion (dark ellipse); (b) geometry of the first experiment in which mesoscopic resistance oscillations were observed [11]; contact areas are hatched.

ity of a microscopic region leads to a concentration of the current, which ensures a large contribution of the oscillating part of its resistance to the total signal at potential contacts (Fig. 3a). The large conductivity of the microscopic region in this case plays the same role as the infinitely large differential conductivity of Shapiro steps in the zeroth approximation. In both cases, the current concentration exceeds the contribution of coherent effects to the signal being measured.

Beginning from the publication by Sharvin [11], who was the first to observe the Aharonov–Bohm effect in a metal film, three generations of interference mesoscopic experiments have been carried out in the following sequence: from a nontrivial experimental geometry [11] to planar submicrometer Au rings [12] and then to GaAs/Al_xGa_{1-x}As heterojunctions in the form of rings [13] prepared strictly according to the rules of nanotechnology. The geometry of the Sharvin experiment was indeed nontrivial: he observed resistance oscillations not on a metal ring, but on a metal film in the form of a cylinder deposited on the lateral surface of a quartz filament. The resistance was measured between points at the vertices of the diagonal of cylinder cross section (Fig. 3b). The total conductance is the sum of two parts: the component oscillating in a magnetic field applied along the cylinder axis and the field-independent component:

$$\frac{1}{R} = \frac{1}{R_0} + \frac{1}{R_{\rm osc}(B)}.$$

It was shown [11] that the amplitude of resistance oscillations in such a system must be on the order of the quantum standard

$$\Delta R_{\rm osc}(B) \sim \frac{2\pi\hbar}{e^2} = 25813\ \Omega$$

(Klitzing constant). Consequently, the field-independent part should be reduced to the maximal possible extent for the constant conductance not to suppress the interference effect; for this reason, the resistance in the Sharvin geometry is not measured along the shortest path.

5. RESISTANCE OSCILLATIONS IN BISMUTH FILMS (EXPERIMENTAL RESULTS)

The resistance of Bi films in our experiments was measured in fields up to 1 T at $T = 4.2$ K. To increase the signal-to-noise ratio, data were accumulated in the temporary memory of a computer and were averaged over six paths to suppress $1/f$ noise. Figure 4 shows the magnetic field dependence of the oscillating part of the resistance; beats correspond to three sinusoids with

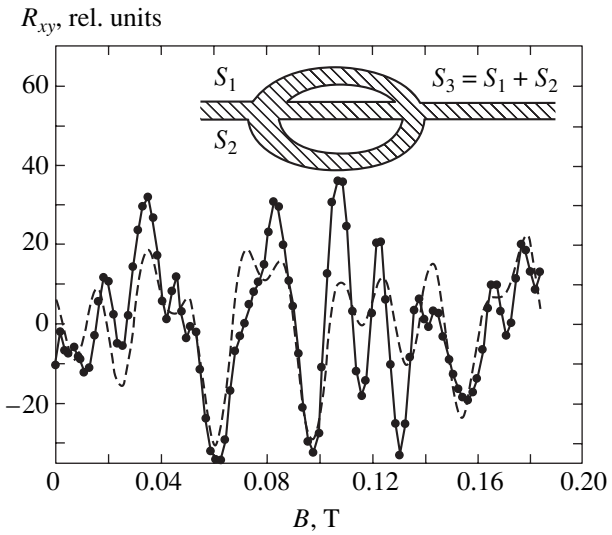


Fig. 4. Mesoscopic resistance oscillations as a function of the applied magnetic field. The ring shown schematically on top has a topology ensuring a ratio of periods $\Delta B_1 = 0.30$ T $= \Phi_0/S_1$, $\Delta B_2 = 0.27$ T $= \Phi_0/S_2$; $\Delta B_3 = 0.11$ T $= \Phi_0/S_3$, $\Phi_0 = \hbar/2e = 2.07 \times 10^{-15}$ Wb, S_1, S_2, and S_3 are the effective areas of different parts of the ring.

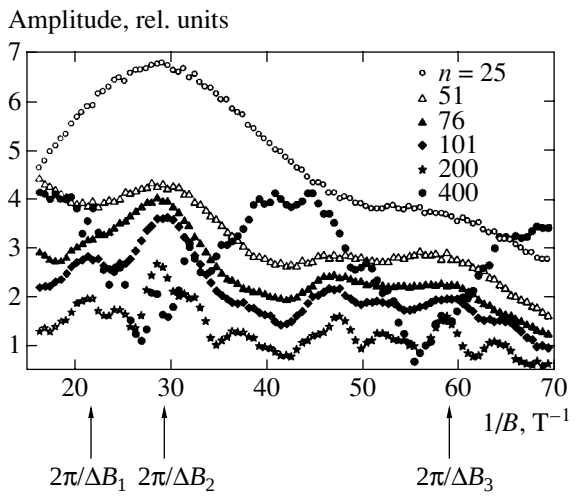


Fig. 5. Fourier spectra of mesoscopic resistance oscillations of the Bi film: n is the number of "measuring points" used for approximate calculation of the Fourier integral; n is proportional to the coherence interval.

periods $\Delta B_1/2\pi = 0.047$ T, $\Delta B_2/2\pi = 0.036$ T, and $\Delta B_3/2\pi = 0.017$ T. The dependence simulating the data (dashed curve) was chosen by the least-squares method using 12 parameters:

$$\Delta R_{xy} = A_1 \sin\left(2\pi\frac{B}{\Delta B_1} + \varphi_1(B)\right) + A_2 \sin\left(2\pi\frac{B}{\Delta B_2} + \varphi_2(B)\right) + A_3 \sin\left(2\pi\frac{B}{\Delta B_3} + \varphi_3(B)\right). \quad (1)$$

Attempts at a standard expansion of experimental dependences into a Fourier series did not allow us to observe spectral peaks revealed by the above-mentioned variational procedure. This is due to the fact that the phase of the oscillation drifts (see Fig. 4) and the Fourier components calculated by the conventional method do not exhibit the typical frequency dependence. We proposed another algorithm allowing for the smallness of the coherence interval in magnetic field. This algorithm substantially simulates in digital form the operation of a two-channel synchronous lock-in detector in the selective mode. The Fourier integral of the product of the function under investigation by sin/cos is taken not over the entire range of the argument, but over the coherence interval. The integration domain in this case gradually moves over the entire range of the argument. Then the obtained dependence of both Fourier coefficients is averaged over the norm. Our algorithm has made it possible to reveal the spectral peaks coinciding (to within 2%) with the periods obtained from the variational procedure (Fig. 5, $n = 101$). The computer time in this case was 10 h, while the time required for obtaining an approximate Fourier expansion was on the order of several minutes. If we take the entire range of the argument as the coherence interval, the procedure described above is transformed into the conventional (exact) Fourier expansion. In this limit, the spectrum of our data becomes irregular and noise-like (Fig. 5, $n = 400$).

The resistance oscillations in our experiment were clearly observed on the Hall (transverse) contacts of the samples. The smallness of the interference contribution to the longitudinal resistance is apparently associated with shunting of the oscillating part by the intrinsic high conductivity of the microscopic region. The transverse resistance of this region under the conditions of the integer quantum Hall effect is high (on the order of the quantum standard $2\pi\hbar/e^2 = 25813\ \Omega$); for this reason, the effect of self-shunting across the Hall contacts is absent. This circumstance is equivalent to the gain obtained in the geometry of the Sharvin experiment. Self-shunting across transverse contacts is suppressed in an increasing field since the transverse Hall resistance of the samples increases in proportion to the field: $R_{xy} = U_y/I_x \propto B$. Consequently, we could expect that the amplitude of oscillations would increase with the field. However, no such increase is observed; a more adequate model of the microscopic region is probably not a ring as an analog of a DC-SQUID, but the microscopic region as a whole as an analog of a solitary Josephson junction with broad banks (a ring without a hole). It is well known that the amplitude of oscillations of the critical Josephson current through a 2D junction of rectangular cross section behaves as

$$I_c^{(J)} \propto \frac{\sin(\pi\Phi/\Phi_0^{(2e)})}{\pi\Phi/\Phi_0^{(2e)}}.$$

In this case, an analogous interference suppression of the amplitude of resistance oscillations,

$$\delta R \propto \frac{\sin(\pi\Phi/\Phi_0^{(ne)})}{\pi\Phi/\Phi_0^{(ne)}},$$

is probably compensated by the suppression of self-shunting effect in an increasing field. Here, $\Phi_0^{(ne)}$ is the magnetic flux quantum for n-electron states.

6. JOSEPHSON EFFECT AND QUANTIZATION OF THE RESISTANCE OF MESOSCOPIC OBJECTS

The above-mentioned analogy of mesoscopic oscillations in the microscopic region with oscillations of the critical current in a DC-SQUID is considerably simplified; however, this analogy correctly reflects, on the whole, the interference origin of these oscillations (in other words, the interference nature of the Aharonov–Bohm effect and the stationary Josephson effect). Expanding this analogy, we should seek an analog of the nonstationary Josephson effect in the microscopic region in question. This effect should be as follows: we

expose a Bi film to microwave radiation and observe, instead of Shapiro steps, periodic deviations from linearity (nonohmicity) of the IV characteristic.

However, it was found in the course of preparation of such an experiment that there exists a simpler experiment revealing the Josephson nature of the microscopic region. Let us consider an extended 1D region (filament) with a phase coherence length exceeding its geometrical length. Then the potential difference applied at the ends of this filament will induce the oscillation of the phase in time with the "Josephson" frequency $\omega_\varphi = eU/\hbar$. This gives a period $T_\varphi = 2\pi\hbar/eU$. On the other hand, the time interval during which an electron traverses the filament is determined by the current, $T^{(1e)} = e/I$. In the case of the equality $T^{(1e)} = T_\varphi$, we obtain a sort of parametric resonance corresponding to the effect of trapping of the filament resistance at the level of the quantum standard,

$$\frac{e}{I} = T^{(1e)} = T_\varphi = \frac{2\pi\hbar}{eU}.$$

Consequently, we have

$$R^{(\text{wire})} = \frac{U}{I} = \frac{2\pi\hbar}{e^2}.$$

Being parametric, the resonance must be repeated for n electrons ($n = 1, 2, 3, \ldots$) passing over a period of phase variation. Then, a smooth variation in the resistance of the microscopic region upon film deposition would give rise to the formation of recurring steps ("sealing") of conductivity with a period multiple of $e^2/2\pi\hbar$. This was precisely observed after a simple statistical processing of the experimental results.

7. QUANTIZATION OF THE RESISTANCE OF BISMUTH FILMS (EXPERIMENTAL RESULTS)

Figure 6 shows correction $\Delta\sigma(L)$ to the conductivity of a Bi film as a function of conductivity, obtained by subtracting the smoothed curve $\langle\sigma(L)\rangle$ from the experimental dependence $\sigma(L)$ (L is the film thickness). The $\Delta\sigma(\sigma_0)$ curve displays oscillations with a period equal on average to $2e^2/2\pi\hbar$.

According to Schwarzchild [12], the period of mesoscopic resistance oscillations also corresponds to a two-electron quantum, $\Phi_0^{(2e)} = 2\pi\hbar/2e$, which corresponds to a second-order magnetodependent contribution to backward scattering. This means that the quantization of the resistance of the mesoscopic region in a magnetic field is of the same origin as the quantization of its conductivity during the film growth. Similar effects of conductivity quantization were also observed earlier for objects of completely different nature (dependence of the resistance on the position of the pin in a tunnel microscope [14] or the behavior of the conductivity of a AsGa channel in a high electron mobility transistor (HEMT) as a function of the voltage across its gate [13]). However, no clear unified conception that would explain the physics of the observed effects was proposed in these publications.

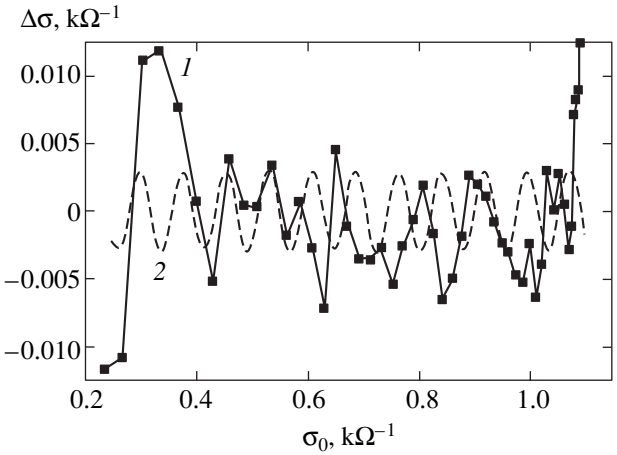


Fig. 6. Conductivity deviation $\Delta\sigma(L) = \sigma(L) - \sigma_0(L)$ from the averaged linear dependence, where $\sigma_0(L) \propto L$. Curve *1* describes the dependence $\Delta\sigma = \Delta\sigma(\sigma_0)$. Curve *2* corresponds to the exact formula of oscillations with a period corresponding to two-electron quantum standard of conductivity $\sin(2\pi\hbar\sigma_0/2e^2)$.

8. CONCLUSIONS

In three parts of this article, we considered coherent effects discovered during electrophysical measurements in disordered systems of three types: the Josephson behavior of an HTSC polycrystal BaKBiO in the phase-separation state (Sections 2 and 3), resistance oscillations in a Bi film in a magnetic field (Sections 4 and 5), and mesoscopic resistance oscillations in the coarse of film growth (Sections 6 and 7).

The general property of such dissimilar systems is the existence of a "synchronizing mechanism" that adjusts the disordered systems to macroscopic interference phenomena.

It was shown in Section 2 that such a synchronizing mechanism, which makes it possible to observe the nonstationary Josephson effect in BaKBiO polycrystals below the phase-separation temperature, can be approximately described as follows. Shapiro steps on their vertical segments possess a high differential conductivity (infinitely high in the simplest model), i.e., a sort of superconducting segments of the IV characteristic. Any current increment smaller than the step height will be short-circuited by a Josephson junction in the randomized 3D circuit, whose working point on the IV curve lies on the vertical segment of a step. In this case, the current increment is distributed over the trajectories corresponding to the maximal differential conductivi-

ties, and the working point on the IV curve lies on the vertical segments of the Shapiro steps for most junctions irrespective of the spread in their critical current. This makes it possible to observe the nonstationary Josephson effect by the method of frequency modulation, although this effect cannot be directly observed on the IV curve of the entire sample due to a considerable spread in the critical currents of Josephson junctions.

The high conductivity of the microscopic region plays the role of an enhancing mechanism that ensures the oscillating correction to the resistance of the Bi film in a magnetic field. On the one hand, the high conductivity in this case causes the concentration of current, which ensures the large contribution of the oscillating part of the microscopic region resistance to the total signal at potential contacts. On the other hand, the high conductivity corresponds to the film region in which the mean free path is large and where the interference effects of the type of the Aharonov–Bohm effect can only be observed. The high conductivity of the microscopic region plays the same role as the infinitely large (in the zeroth approximation) differential conductivity of Shapiro steps. In both cases, the current concentration increases the contribution of coherent effects to the signal being measured.

In the third case, the synchronizing mechanism for the oscillating correction to the conductivity, whose periodicity corresponds to resistance quantization, is apparently the local nonuniformity in the rate of film growth during laser sputtering. In the case of slow thermal deposition or laser sputtering alternating with film annealing, such an effect should not be observed. In our experiments, this effect was observed only in some series of laser sputtering. The effect is very sensitive to the growth rate and depends on the laser pulse energy. The quality of prepared substrate surface also plays a certain role.

In the coarse of nonuniform growth, extended percolation channels are obviously formed in the film; the resistance of these channels ensures a quantizing contribution to the total conductivity of the sample. The fact that the channels are threadlike, narrow, extended objects is indirectly confirmed by the small amplitude of the Hall resistance oscillations [9, 15] in quantizing magnetic fields. Indeed, according to the semiclassical theory [16], the Hall voltage turns out to be equal to the potential difference across the open edge trajectories of carriers. Thus, the larger the width of a percolation channel, the higher the difference in the chemical potentials across its ends and the higher the Hall emf. Accordingly, the percolation in narrow channels corresponds to a low Hall voltage and a low Hall resistance.

ACKNOWLEDGMENTS

The authors are grateful to A.P. Rusakov for providing polycrystalline $Ba_{0.6}K_{0.4}BiO_3$ samples.

This study was supported financially by the Russian Foundation for Basic Research (project no. 01-02-17981), the II-3 program "Strongly Correlated Electrons in Semiconductors, Metals, Superconductors and Magnetic Materials" of the Physics Division of the Russian Academy of Sciences, and the Federal Target Scientific and Engineering Program "Experimental and Theoretical Studies of HTSC" (state contract no. 40.012.1.1.1357).

REFERENCES

1. B. D. Josephson, Phys. Lett. **1**, 251 (1962).
2. S. Shapiro, A. R. Janus, and S. Holly, Rev. Mod. Phys. **36**, 223 (1964).
3. N. Kanter and F. L. Vernon, J. Appl. Phys. **43**, 3174 (1972).
4. N. V. Anshukova, V. B. Ginodman, A. I. Golovashkin, *et al.*, Zh. Éksp. Teor. Fiz. **97**, 1635 (1990) [Sov. Phys. JETP **70**, 923 (1990)].
5. L. N. Zherikhina, A. I. Golovashkin, A. V. Gudenko, and A. M. Tskhovrebov, Czech. J. Phys., Suppl. **46**, 851 (1996).
6. A. A. Gorbatsevich, Yu. V. Kopaev, and I. V. Tokatly, Pis'ma Zh. Éksp. Teor. Fiz. **52**, 736 (1990) [JETP Lett. **52**, 95 (1990)].
7. B. L. Al'tshuler, A. G. Aronov, and B. Z. Spivak, Pis'ma Zh. Éksp. Teor. Fiz. **33**, 101 (1981) [JETP Lett. **33**, 94 (1981)].
8. B. L. Altshuler and P. A. Lee, Phys. Today **41** (12), 37 (1988).
9. A. N. Zherikhin, L. N. Zherikhina, G. V. Kuleshova, *et al.*, in *Proceedings of Scientific Session of Moscow Institute of Engineering Physics, MIFI-2001* (Mosk. Inzh.–Fiz. Inst., Moscow, 2001), Vol. 4, p. 110.
10. A. N. Zherikhin, G. Yu. Shubnyĭ, L. N. Zherikhina, *et al.*, Poverkhnost, No. 6, 79 (2000).
11. D. Yu. Sharvin and Yu. V. Sharvin, Pis'ma Zh. Éksp. Teor. Fiz. **34**, 285 (1981) [JETP Lett. **34**, 272 (1981)].
12. B. Schwarzchild, Phys. Today **39**, 17 (1986).
13. A. E. Hansem, cond-mat/9909246.
14. J. L. Costa-Kromer, Phys. Rev. Lett. **78**, 4990 (1997).
15. A. I. Golovashkin, A. N. Zherikhin, L. N. Zherikhina, *et al.*, in *Proceedings of Scientific Session of Moscow Institute of Engineering Physics, MIFI-2003* (Mosk. Inzh.–Fiz. Inst., Moscow, 2003), Vol. 4, p. 108.
16. S. M. Girvin, *The Quantum Hall Effect: Lectures Delivered at Ecole d'Ete Les Houches* (Indiana Univ. Press, Bloomington, 1998).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 99, No. 2, 2004, pp. 379–385.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 2, 2004, pp. 426–434.

SOLIDS
Electronic Properties

Monopolar Optical Orientation of Electron Spins in Bulk Semiconductors and Heterostructures

E. L. Ivchenko and S. A. Tarasenko

Ioffe Physicotechnical Institute, Russian Academy of Sciences, St. Petersburg, 194021 Russia
e-mail: ivchenko@coherent.ioffe.ru

Received February 9, 2004

Abstract—It is shown that intraband absorption of circularly polarized light leads to spin polarization of the electron gas. A theory of this monopolar optical spin orientation is developed for indirect intraband transitions in bulk semiconductors and for indirect intrasubband and direct intersubband transitions in quantum wells. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Spin-dependent phenomena in semiconductor structures are the subject of extensive ongoing research. The most widespread method for creating spin polarization and investigating kinetics of spin-polarized carriers is optical orientation of electron spins [1]. In interband excitation by circularly polarized light, direct transitions from the valence band to the conduction band can occur only with electron angular momentum changing by ±1. This selection rule controls spin orientation of photoexcited carriers. Both direction and degree of spin orientation depend on the direction of light propagation and the degree of circular polarization.

To date, optical orientation has been studied in detail for interband optical absorption in bulk semiconductors and nanostructures (see [1–4]) and, to some extent, for direct intersubband transitions in the complex valence bands Γ_8 of semiconductors with zinc-blende-type lattices [3, 5, 6]. Monopolar optical orientation of the electron gas was not addressed until recently. However, the circular photogalvanic and spin-galvanic effects discovered in *n*-type InAs/AlGaAs and GaAs/AlGaAs quantum wells in [7–12] indicate that intraband absorption of circularly polarized light results in redistribution of electrons between spin states.

In this paper, we present a theory of monopolar optical orientation induced in the electron gas by circularly polarized light with a photon energy much smaller than the band gap. We consider indirect intraband optical transitions in bulk semiconductors and both indirect intrasubband and direct intersubband transitions in quantum wells based on semiconductors with zinc-blende-type lattices. Since only electrons are involved in photoexcitation, monopolar orientation can be considered as an optical spin-generation method.

The paper is organized as follows. In Section 2, we consider monopolar optical orientation of the electron gas in bulk semiconductors. Section 3 presents a theory of optical spin orientation of electrons in quantum wells under intra- and intersubband excitation. In Section 4, we consider the short-range Elliott–Yafet spin relaxation mechanism, which is determined by the same constants of interband scattering by phonons and defects that determine electron spin orientation by indirect optical transitions. In the Conclusions, calculated results are compared with available experimental data.

2. OPTICAL ORIENTATION IN BULK SEMICONDUCTORS

Intraband (Drude) absorption of light must involve electron momentum transfer, because energy and momentum must be conserved simultaneously. The role of a third particle that transfers momentum to an electron can be played by a static defect, an acoustic or optical phonon, or another electron.

In perturbation theory, indirect optical transitions are treated as second-order processes involving virtual intermediate states. The composite matrix element for a transition between states in the conduction band is written in the following standard form [3]:

$$M_{s'\mathbf{k}' \leftarrow s\mathbf{k}} = \sum_{\nu}\left(\frac{V_{cs'\mathbf{k}', \nu\mathbf{k}} R_{\nu\mathbf{k}, cs\mathbf{k}}}{E_{c\mathbf{k}} - E_{\nu\mathbf{k}} + \hbar\omega} + \frac{R_{cs'\mathbf{k}', \nu\mathbf{k}'} V_{\nu\mathbf{k}', cs\mathbf{k}}}{E_{c\mathbf{k}} - E_{\nu\mathbf{k}'} \mp \hbar\Omega_{\mathbf{k}-\mathbf{k}'}}\right), \quad (1)$$

where $\mathbf{k}$ and $\mathbf{k}'$ are the wave vectors of an electron in the initial and final states, respectively; s is the spin index; ν refers to the band of an intermediate state; $E_{c\mathbf{k}}$, $E_{c\mathbf{k}'}$, and E_ν are the electron energies in the initial, final, and intermediate states, respectively; R is a matrix element of the electron–electromagnetic-wave interaction operator; V is a matrix element of the operator of electron–phonon interaction or electron scattering by an impurity; $\hbar\omega$ is a quantum of light; and $\hbar\Omega_{\mathbf{k}-\mathbf{k}'}$ is the phonon

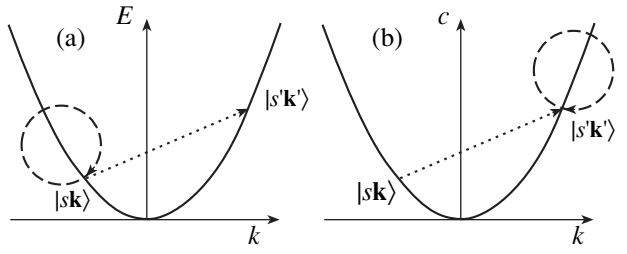


Fig. 1. Indirect intraband optical transitions. Dashed and dotted curves represent electron–electromagnetic-wave interaction and electron scattering, respectively. Panels a and b correspond to the first and second terms in composite matrix element (1), respectively.

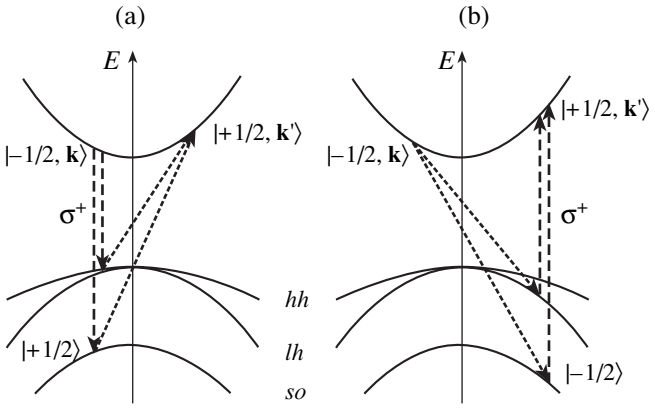


Fig. 2. Optical orientation by indirect optical transitions in the conduction band with intermediate states in the valence band.

energy ($\hbar\Omega = 0$ for elastic scattering by an impurity). The minus and plus signs in $\mp$ correspond to phonon emission and absorption, respectively. The matrix element R for the absorptive transition from $|n, s, \mathbf{k}\rangle$ to $|\nu, \mathbf{k}\rangle$ is expressed in terms of the corresponding matrix element of the momentum operator $\hat{\mathbf{p}}$ as

$$R_{\nu\mathbf{k},\, ns\mathbf{k}} = \frac{e}{cm_0}\mathbf{A}\cdot\mathbf{p}_{\nu\mathbf{k},\, ns\mathbf{k}}, \tag{2}$$

where m_0 is the free-electron mass, e is the elementary charge, $\mathbf{A} = A\mathbf{e}$ is the complex amplitude of electromagnetic vector potential ($A = |\mathbf{A}|$), and $\mathbf{e}$ is the polarization vector.

Intraband absorption is dominated by transitions via intermediate states in the same band, as schematized in Fig. 1. The corresponding transition matrix element can readily be derived from (1) by assuming that the intermediate state is in the conduction band. When the photon energy is substantially larger than the energy transferred to the electron in a scattering event ($\hbar\omega \gg \hbar\Omega_{\mathbf{k}-\mathbf{k}'}$), the matrix element can be written as

$$M^{(0)}_{cs'\mathbf{k}'\leftarrow cs\mathbf{k}} = \frac{eA}{c\omega\hbar^2}\mathbf{e}\left(\frac{dE_{c\mathbf{k}}}{d\mathbf{k}} - \frac{dE_{c\mathbf{k}'}}{d\mathbf{k}'}\right)V_{cs'\mathbf{k}',\, cs\mathbf{k}}. \tag{3}$$

Processes of this kind determine the intraband absorption coefficient, but do not contribute to optical orientation, because the absolute value of (3) is independent of the sign of circular polarization.

Spin orientation caused by intraband absorption of circularly polarized light can be modeled by analyzing virtual transitions via intermediate states in the complex valence band and taking into account spin-orbit splitting. Figure 2 is a schematic illustration of spin orientation of the electron gas. Excitation by circularly polarized light σ^+ induces electron $|-1/2\rangle \longrightarrow |+1/2\rangle$ spin-flip transitions via intermediate states in the light-hole and spin-orbit split-off subbands, whereas reverse ($|+1/2\rangle \longrightarrow |-1/2\rangle$) transitions are forbidden. This photoexcitation asymmetry is responsible for spin orientation of the electron gas.

Let us consider acoustic-phonon-assisted optical orientation. The Hamiltonian term that represents the interband mixing due to the acoustic-phonon deformation potential is a 2-by-6 matrix:

$$\hat{V}_{\rm ph} = \Xi_{\rm cv}\begin{bmatrix} -\dfrac{\varepsilon_{yz}+i\varepsilon_{xz}}{\sqrt{2}} & \sqrt{\dfrac{2}{3}}\varepsilon_{xy} & \dfrac{\varepsilon_{yz}-i\varepsilon_{xz}}{\sqrt{6}} & 0 & -\dfrac{\varepsilon_{xy}}{\sqrt{3}} & -\dfrac{\varepsilon_{yz}-i\varepsilon_{xz}}{\sqrt{3}} \\ 0 & -\dfrac{\varepsilon_{yz}+i\varepsilon_{xz}}{\sqrt{6}} & \sqrt{\dfrac{2}{3}}\varepsilon_{xy} & \dfrac{\varepsilon_{yz}-i\varepsilon_{xz}}{\sqrt{2}} & -\dfrac{\varepsilon_{yz}+i\varepsilon_{xz}}{\sqrt{3}} & \dfrac{1}{\sqrt{3}}\varepsilon_{xy} \end{bmatrix}, \tag{4}$$

where $\Xi_{\rm cv}$ denotes the interband deformation-potential constant; ε_{yz}, ε_{xz}, and ε_{xy} are the off-diagonal components of the strain tensor, which are transformed as a vector under the T_d group; and both the Γ_6 conduction-band Bloch functions and the Γ_8 and Γ_7 valence-band Bloch functions are defined in the canonical basis [3].

Taking into account virtual intermediate states in the valence band, we write the matrix element for an indirect optical transition as

$$M_{s'\mathbf{k}',\, s\mathbf{k}} = \mathscr{A}_{\mathbf{k}',\,\mathbf{k}}\delta_{s's} + \mathscr{B}_{\mathbf{k}',\,\mathbf{k}}\cdot\boldsymbol{\sigma}_{s's}, \tag{5}$$

with $\mathscr{A}_{\mathbf{k}',\,\mathbf{k}} \equiv M^{(0)}$ and $\mathscr{B}_{\mathbf{k}',\,\mathbf{k}}$ representing the contribu-

tions of intermediate states in the conduction and valence bands. Here,

$$\mathcal{B}_{\mathbf{k}',\mathbf{k}} = \frac{2}{3}\frac{eA}{cm_0}\frac{\Delta_{so}}{E_g(E_g+\Delta_{so})}iP_{cv}[\mathbf{U}_{\mathbf{k}',\mathbf{k}}\times\mathbf{e}], \quad (6)$$

E_g is the band gap; Δ_{so} is the valence-band spin-orbit splitting; $P_{cv} = \langle S|\hat{p}_z|Z\rangle$ is the interband momentum matrix element; σ_x, σ_y, and σ_z are the Pauli matrices; and the $\mathbf{U}_{\mathbf{k}',\mathbf{k}}$ vector components are the off-diagonal-strain matrix elements

$$\mathbf{U}_{\mathbf{k}',\mathbf{k}} = \Xi_{cv}\begin{pmatrix}\langle\mathbf{k}'|\varepsilon_{yz}|\mathbf{k}\rangle\\ \langle\mathbf{k}'|\varepsilon_{xz}|\mathbf{k}\rangle\\ \langle\mathbf{k}'|\varepsilon_{xy}|\mathbf{k}\rangle\end{pmatrix}$$

calculated in a plane-wave basis. According to (6), the spin-dependent contribution to the intraband optical-transition matrix element, $\mathcal{B}_{\mathbf{k}',\mathbf{k}}$, is proportional to the valence-band spin-orbit splitting. In mathematical terms, this is explained by the opposite signs of the matrix elements corresponding to virtual transitions via the heavy- or light-hole Γ_8 subbands and the spin-orbit split-off Γ_7 subband.

In the expression for spin generation rate,

$$\dot{\mathbf{S}} = \sum_{\mathbf{k}}\mathrm{Tr}\left[\frac{1}{2}\hat{\boldsymbol{\sigma}}\dot{\rho}(\mathbf{k})\right], \quad (7)$$

the matrix $\dot{\rho}(\mathbf{k})$ is expressed in the linear approximation with respect to light intensity as follows (see [13]):

$$\dot{\rho}_{ss'}(\mathbf{k}) = \frac{2\pi}{\hbar}\sum_{\mathbf{k}_1,s_1}(f^0_{\mathbf{k}_1}-f^0_{\mathbf{k}})$$
$$\times[M_{s\mathbf{k},s_1\mathbf{k}_1}M^*_{s'\mathbf{k},s_1\mathbf{k}_1}\delta(E_{\mathbf{k}_1}-E_{\mathbf{k}}+\hbar\omega) \quad (8)$$
$$+ M_{s_1\mathbf{k}_1,s'\mathbf{k}}M^*_{s_1\mathbf{k}_1,s\mathbf{k}}\delta(E_{\mathbf{k}_1}-E_{\mathbf{k}}-\hbar\omega)],$$

where $f^0_{\mathbf{k}}$ denotes the equilibrium carrier distribution function. Performing summation over $\mathbf{k}$ and $\mathbf{k}_1$ and using the approximate relation

$$\frac{1}{m^*}\approx\frac{2}{3}\frac{|P_{cv}|^2}{m_0^2}\left(\frac{2}{E_g}+\frac{1}{E_g+\Delta_{so}}\right)$$

for the effective electron mass, we obtain a final expression for the rate of spin generation by intraband absorption concurrent with acoustic-phonon-assisted electron scattering:

$$\dot{\mathbf{S}}_{ph} = \frac{1}{6}\frac{\Xi_{cv}^2}{\Xi_c^2}\frac{\Delta_{so}^2}{E_g(E_g+\Delta_{so})(3E_g+2\Delta_{so})}$$
$$\times\left(\frac{2}{5}+\frac{3}{5}\frac{v_L^2}{v_T^2}\right)IK_{ph}P_{circ}\mathbf{l}. \quad (9)$$

Here, Ξ_c is the intraband deformation potential constant; v_L and v_T denote the longitudinal and transverse sound velocities, respectively; $I = A^2\omega^2 n_\omega/2\pi c$ is light intensity; and P_{circ} is the degree of circular polarization related to the unit vector $\mathbf{l}$ in the direction of light propagation as $P_{circ}\mathbf{l} = i[\mathbf{e}\times\mathbf{e}^*]$. The optical absorption coefficient K_{ph} is

$$K_{ph} = \frac{4\alpha}{3n_\omega}\left(\frac{\Xi_c}{\hbar\omega}\right)^2\frac{k_BT}{\rho v_L^2}\left(\frac{2m^*\omega}{\hbar}\right)^{1/2}N_e,$$

where $\alpha = e^2/\hbar c \approx 1/137$ is the fine-structure constant, n_ω is refractive index, k_B is the Boltzmann constant, T is temperature, ρ is the crystal density, and N_e is the electron concentration.

At low temperatures, electron scattering by static defects may predominate over electron–phonon interaction. Defect-assisted intrasubband absorption of circularly polarized light can result in spin orientation if the scattering defects cause mixing of valence- and conduction-band states. For example, the defects may be deep impurities with p Bloch states. In the case of short-range defects, the interband-scattering term of the Hamiltonian has the form

$$\hat{V}_{def} = \begin{bmatrix}-\frac{V_x+iV_y}{\sqrt{2}} & \sqrt{\frac{2}{3}}V_z & \frac{V_x-iV_y}{\sqrt{6}} & 0 & -\frac{V_z}{\sqrt{3}} & -\frac{V_x-iV_y}{\sqrt{3}}\\ 0 & -\frac{V_x+iV_y}{\sqrt{6}} & \sqrt{\frac{2}{3}}V_z & \frac{V_x-iV_y}{\sqrt{2}} & -\frac{V_x+iV_y}{\sqrt{3}} & \frac{V_z}{\sqrt{3}}\end{bmatrix}\sum_j\delta(\mathbf{r}-\mathbf{r}_j), \quad (10)$$

where V_x, V_y, and V_z are the matrix elements associated with defect-induced mixing of the conduction-band Bloch function S with the valence-band functions X, Y, and Z; $\mathbf{r}_j$ is an impurity site. The spin-dependent contribution to the photon absorption matrix element, which is determined by virtual transitions via intermediate

states in the valence band and is responsible for spin orientation, is given by (6) with

$$\mathbf{U}_{\mathbf{k}',\mathbf{k}} = \begin{pmatrix} V_x \\ V_y \\ V_z \end{pmatrix} \sum_j \exp[i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}_j]. \qquad (11)$$

Averaging over $\mathbf{r}_j$ and impurity wavefunctions ($\langle V_\alpha V_\alpha\rangle \equiv V_{cv}^2$ and $\langle V_\alpha V_\beta\rangle = 0$ if $\alpha \neq \beta$), we obtain a final expression for the rate of spin generation due to static-defect-assisted intraband optical transitions:

$$\dot{\mathbf{S}}_{def} = \frac{V_{cv}^2}{V_c^2} \times \frac{\Delta_{so}^2}{E_g(E_g+\Delta_{so})(3E_g+2\Delta_{so})} IK_{def}P_{circ}\mathbf{l}, \qquad (12)$$

where V_c is the matrix element for intraband electron scattering by defects. The optical absorption coefficient K_{def} is expressed as

$$K_{def} = \frac{4\alpha}{3n_\omega}\frac{V_c^2 N_{def}}{(\hbar\omega)^2}\left(\frac{2m^*\omega}{\hbar}\right)^{1/2} N_e,$$

where N_{def} denotes the defect concentration.

Using (9) with the parameters and constants $E_g \approx 1.5$ eV, $\Delta_{so} \approx 0.34$ eV, $\Xi_c \approx -8.3$ eV, and $\Xi_{cv} \approx 3$ eV characteristic of GaAs (see [1, Ch. 3] and [14]), we estimate the average spin generated by acoustic-phonon-assisted intraband absorption of a circularly polarized photon of energy $\hbar\omega = 10$ meV as 3×10^{-6}. A higher value would be obtained if other scattering mechanisms were taken into account.

3. OPTICAL ORIENTATION IN QUANTUM WELLS

In quantum-well structures, intraband absorption can be attributed to both indirect intrasubband optical transitions and direct intersubband transitions between size-quantized subbands (if photon energy corresponds to energy separation between subbands).

3.1. Intrasubband Transitions

As in the case of monopolar optical orientation by intraband transitions in bulk semiconductors, indirect intrasubband optical transitions resulting in monopolar optical orientation in quantum wells involve virtual intermediate states in the complex valence band.

We assume here that electrons occupy the ground size-quantized subband ($e1$) in an infinitely deep rectangular quantum well and the size-quantization energy is substantially larger than the mean kinetic energy in the (x, y) plane. In this case, virtual transitions via the valence band must involve intermediate states in the ground heavy-hole, light-hole, and spin-orbit split-off subbands.

In the case of intrasubband optical absorption concurrent with electron scattering by bulk acoustic phonons, the spin generation rate is

$$\dot{\mathbf{S}}_{ph} = \frac{1}{6}\frac{\Xi_{cv}^2}{\Xi_c^2}\frac{\Delta_{so}^2}{E_g(E_g+\Delta_{so})(3E_g+2\Delta_{so})} \times \left[\mathbf{l}_\parallel + \frac{a}{3}\sqrt{\frac{2m^*\omega}{\hbar}}\left(\frac{1}{8}+\frac{7}{8}\frac{v_L^2}{v_T^2}\right)\mathbf{l}_z\right] IP_{circ}\eta_{ph}, \qquad (13)$$

where $\mathbf{l} = (\mathbf{l}_\parallel, \mathbf{l}_z)$ is the unit vector in the direction of light propagation ($\mathbf{l}_\parallel$ and $\mathbf{l}_z$ are its components parallel and perpendicular to the boundary), the fraction of energy flux η_{ph} absorbed in the quantum well under normal incidence is expressed as

$$\eta_{ph} = \frac{3\pi\alpha}{n_\omega}\frac{\Xi_c^2}{(\hbar\omega)^2}\frac{k_B T}{\rho a v_L^2} N_e;$$

a is the quantum-well width, and N_e is the two-dimensional electron concentration.

For light absorption concurrent with elastic scattering by defects, the spin generation rate is

$$\dot{\mathbf{S}}_{def} = \frac{2}{3}\frac{\Delta_{so}^2}{E_g(E_g+\Delta_{so})(3E_g+2\Delta_{so})} \times \left(\frac{V_\parallel^2}{V_c^2}\mathbf{l}_\parallel + \frac{V_z^2}{V_c^2}\mathbf{l}_z\right) IP_{circ}\eta_{def}, \qquad (14)$$

where $V_\parallel = V_x = V_y$ and V_z are the matrix elements for defect-induced mixing of the conduction-band Bloch function S with the valence-band wavefunctions X, Y, and Z, the fraction of energy flux η_{def} absorbed in the quantum well under normal incidence is expressed as

$$\eta_{def} = \frac{2\pi\alpha}{n_\omega}\frac{V_c^2 N_{def}}{(\hbar\omega)^2} N_e,$$

with N_{def} denoting the two-dimensional defect concentration, and V_c is the matrix element for intraband electron scattering by defects.

3.2. Direct Intersubband Transitions

Intersubband light absorption in quantum wells is a resonant process that can occur if the photon energy equals the energy separation between subbands. In the single-band approximation, direct optical transitions from the subband $e1$ to the subband $e2$ conserve spin

orientation and occur only if the polarization vector **e** has a nonzero normal component e_z [3, 15–17].

Monopolar optical orientation of the electron gas by direct intersubband optical transitions can be modeled only if mixing of valence-band states with conduction-band wavefunctions is allowed for. Under this condition, light polarized in the quantum-well plane can be absorbed and intersubband transitions are made spin-dependent by circularly polarized light.

We assume that electrons occupy the ground subband ($e1$) and the size-quantization energy is substantially larger than the mean kinetic energy in the (x, y) plane. To allow for $\mathbf{k}\cdot\mathbf{p}$ mixing, we represent the optical-transition matrix elements as

$$M_{s',s} = \mathcal{A}\delta_{s's} + \mathcal{B}\cdot\boldsymbol{\sigma}_{s's}, \tag{15}$$

where the spin-independent contribution is the matrix element

$$\mathcal{A} = \frac{eA}{cm^*}p_{21}e_z, \tag{16}$$

calculated in the single-band approximation, and the contribution due to $\mathbf{k}\cdot\mathbf{p}$ mixing is represented by

$$\mathcal{B} = -\frac{i}{2}\frac{eA}{cm^*} \times \frac{E_{21}(2E_g+\Delta_{so})\Delta_{so}}{E_g(E_g+\Delta_{so})(3E_g+2\Delta_{so})}[\mathbf{e}\times\mathbf{n}_z]p_{21}. \tag{17}$$

Here, $E_{21} = E_2 - E_1$ is the energy separation between the subbands, $\mathbf{n}_z$ is the unit vector along the z axis, and the matrix element p_{21} of the momentum operator $\hat{p}_z = -i\hbar\partial/\partial z$ is calculated as

$$p_{21} = \int dz u_2(z)\hat{p}_z u_1(z)$$

by using size-quantization wavefunctions.

Intersubband light absorption involves redistribution of carriers between subbands. The corresponding spin density matrices are determined by the equations

$$\dot{\rho}_{ss'}^{(e1)}(\mathbf{k}) = -\frac{2\pi}{\hbar}f_{\mathbf{k}}^0\sum_{s_1}M_{s_1,s'}M_{s_1,s}^*\delta(\hbar\omega - E_{21}),$$
$$\dot{\rho}_{ss'}^{(e2)}(\mathbf{k}) = \frac{2\pi}{\hbar}f_{\mathbf{k}}^0\sum_{s_1}M_{s,s_1}M_{s',s_1}^*\delta(\hbar\omega - E_{21}). \tag{18}$$

Substituting (18) into (7), we obtain a final expression for the spin generation rate in the lower and upper subbands due to absorption of circularly polarized light:

$$\dot{\mathbf{S}}_{(e1/e2)} = \frac{1}{2}(\eta_\parallel\mathbf{l}_z \pm \sqrt{\eta_z\eta_\parallel}\mathbf{l}_\parallel)\frac{I}{\hbar\omega}P_{circ}, \tag{19}$$

where the fraction of energy flux absorbed in the quantum well for light polarized along the z axis is

$$\eta_z = \frac{4\pi^2\alpha\hbar|p_{21}|^2}{n_\omega m^{*2}\omega}N_e\delta(\hbar\omega - E_{21})$$

and the fraction of absorbed energy flux due to $\mathbf{k}\cdot\mathbf{p}$ mixing for light polarized in the quantum-well plane is [16]

$$\eta_\parallel = \frac{1}{4}\frac{E_{21}^2(2E_g+\Delta_{so})^2\Delta_{so}^2}{E_g^2(E_g+\Delta_{so})^2(3E_g+2\Delta_{so})^2}\eta_z.$$

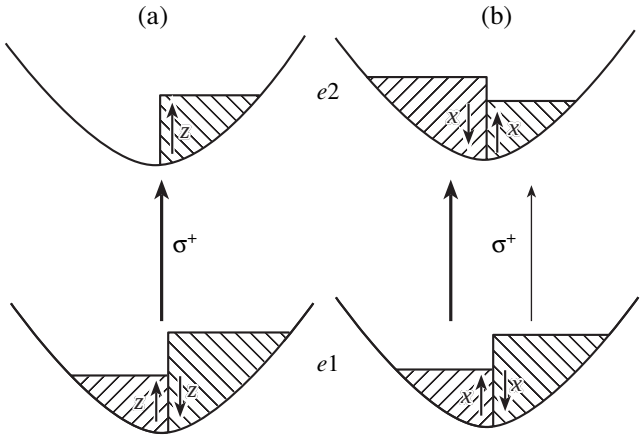


Fig. 3. Intersubband optical orientation of electron spin via (a) spin-flip transitions and (b) spin-conserving transitions.

Expression (19) describes optical orientation of a two-dimensional electron gas by intersubband transitions. Under normal incidence ($l_\parallel = 0$), circularly polarized light induces intersubband spin-flip transitions. Therefore, the electron spins generated in both subbands are parallel (see Fig. 3a). When circularly polarized light is incident at an oblique angle, the polarization component e_z induces fast spin-conserving intersubband transitions. However, their rates for spins oriented parallel and antiparallel to $\mathbf{l}_\parallel$ differ by the factor $(\eta_z\eta_\parallel)^{1/2}$. As a result, the spin parallel to $\mathbf{l}_\parallel$ is redistributed between subbands, while the total spin in the quantum-well plane is conserved in the photoexcitation process (see Fig. 3b). The redistribution of the spin parallel to $\mathbf{l}_\parallel$ between subbands gives rise to an average spin orientation in the quantum-well plane if hot electrons are spin-depolarized in the course of energy relaxation. In this case, the average spin generation rate in the quantum-well plane is $\xi\dot{\mathbf{S}}_{(e1)}$, where ξ is a dimensionless factor introduced to take into account spin depolarization.

4. SPIN RELAXATION VIA INTERBAND SCATTERING

Virtual interband scattering can lead to both monopolar optical orientation and spin relaxation in the electron gas. At the microscopic level, this spin relaxation

mechanism can be explained by $\mathbf{k}\cdot\mathbf{p}$ admixture of valence-band states to the conduction-band wavefunction and virtual interband scattering by phonons or a static defects. For bulk semiconductors, this short-range Elliott–Yafet spin relaxation mechanism was discussed by Pikus and Titkov in [1, Ch. 3].

The spin relaxation time corresponding to this mechanism can be calculated by using composite matrix element (1) with the electron–electromagnetic-wave interaction operator R replaced by the $\mathbf{k}\cdot\mathbf{p}$-mixing operator $(\hbar/m_0)\mathbf{k}\cdot\mathbf{p}_{vk,\,ck}$.

Let us consider electron spin relaxation in an infinitely deep rectangular quantum well. When the transitions via intermediate states in the subbands $hh1$, $lh1$, and $so1$ are taken into account, the spin-flip matrix element is given by expression (6) with $(eA/c)\mathbf{e}$ replaced by $\hbar(\mathbf{k}+\mathbf{k}')/2$, where $\mathbf{k}$ and $\mathbf{k}'$ are wave vectors in the quantum-well plane.

If the two-dimensional electron gas is characterized by the Boltzmann distribution, then the spin relaxation times in the quantum-well plane and along the z axis due to electron scattering by bulk acoustic phonons are

$$\frac{1}{\tau_{s\parallel}} = \frac{1}{3}\frac{\Xi_{cv}^2}{\Xi_c^2}\frac{v_L^2}{v_T^2} \times \frac{\Delta_{so}^2 k_B T}{E_g(E_g+\Delta_{so})(3E_g+2\Delta_{so})}\frac{1}{\tau_{ph}}, \qquad (20)$$

$$\frac{1}{\tau_{sz}} = \frac{2a}{3\sqrt{\pi}}\sqrt{\frac{2m^*}{\hbar^2}k_B T}\left(\frac{23}{35}+\frac{12}{35}\frac{v_T^2}{v_L^2}\right)\frac{1}{\tau_{s\parallel}},$$

where the electron momentum relaxation time τ_{ph} due to electron scattering by acoustic phonons is defined by the relation

$$\frac{1}{\tau_{ph}} = \frac{3}{2}\frac{m^*\Xi_c^2}{\rho a v_L^2 \hbar^3}k_B T.$$

The spin relaxation times due to interband scattering by static short-range defects are

$$\frac{1}{\tau_{sz}} = \frac{4}{3}\frac{V_z^2}{V_c^2}\frac{\Delta_{so}^2 k_B T}{E_g(E_g+\Delta_{so})(3E_g+2\Delta_{so})}\frac{1}{\tau_{def}}, \qquad (21)$$

$$\frac{1}{\tau_{s\parallel}} = \left(\frac{1}{2}+\frac{V_\parallel^2}{V_z^2}\right)\frac{1}{\tau_{sz}},$$

where the electron momentum relaxation time τ_{def} due to short-range defects is defined by the relation

$$\frac{1}{\tau_{def}} = \frac{m^*}{\hbar^3}V_c^2 N_{def}.$$

Note that spin relaxation of a two-dimensional electron gas controlled by the short-range Elliott–Yafet mechanism can be strongly anisotropic.

5. CONCLUSIONS

Before comparing theory with experiment, we note that monopolar optical orientation of electron spin was detected experimentally by observing spin-galvanic and circular photogalvanic effects in n-type quantum wells [7–12]. In the spin-galvanic effect, electric current is generated as a result of spin-flip relaxation of spin-polarized electrons. This effect was observed under conditions of monopolar optical orientation in quantum wells based on GaAs and InGaAs semiconductors. The nonequilibrium electron spin in the quantum-well plane required to observe the spin-galvanic effect in [001]-grown structures was generated either by circularly polarized light under normal incidence and spin precession in external magnetic field [8, 9, 12] or under oblique incidence in the absence of magnetic field [11].

The photocurrent due to direct optical transitions in a GaAs-based structure with 30 8.2-nm-wide quantum wells and a carrier concentration of $N_e \sim 2\times10^{11}$ cm^{-2} in each well was about 1 nA/W for light incident at an angle of 20° at room temperature [11]. Estimating the spin-galvanic current density as $j_{SGE} \propto e\tau_p(\beta/\hbar)\dot{S}$ (see [11]), where $t_p \sim 10^{-13}$ s is the electron momentum relaxation time at room temperature and $\beta/\hbar \sim 10^6$ cm/s is the constant parameter determining the linear-in-momentum spin-orbit splitting of electron states, the experimental rate of spin generation in the plane of each quantum well can be estimated as $\dot{S}/I \sim 10^7$ erg^{-1}. According to the results obtained in Section 3.2, intersubband optical orientation in the quantum-well plane should be attributed to spin-conserving intersubband transitions followed by spin depolarization of hot carriers. A theoretical estimate of $\dot{S}_\parallel/I$ obtained by using (19) and taking into account an inhomogeneous broadening of 10 meV also gives 10^7 erg^{-1}. The rate of spin generation via intrasubband optical transitions determined in experiment on the spin-galvanic effect is higher than the corresponding theoretical value by an order of magnitude. This may be explained by the fact that the optical orientation caused by light absorption via indirect transitions includes an additional contribution due to electron–electron scattering, which does not affect electron–gas mobility, but can strongly modify the absolute value of the intrasubband optical absorption coefficient [18].

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, INTAS, the Russian Academy of Sciences, the Ministry of Industry and Science, and the Dynasty Foundation, which is affiliated with the International Center for Fundamental Physics in Moscow.

REFERENCES

1. *Optical Orientation*, Ed. by F. Meier and B. P. Zakharchenya (North-Holland, Amsterdam, 1984; Nauka, Leningrad, 1989).
2. I. A. Merkulov, V. I. Perel', and M. E. Portnoĭ, Zh. Éksp. Teor. Fiz. **99**, 1202 (1990) [Sov. Phys. JETP **72**, 669 (1990)].
3. E. L. Ivchenko and G. E. Pikus, *Superlattices and Other Heterostructures: Symmetry and Optical Phenomena* (Springer, Berlin, 1995).
4. I. A. Akimov, D. N. Mirlin, V. I. Perel', and V. F. Sapega, Fiz. Tekh. Poluprovodn. (St. Petersburg) **35**, 758 (2001) [Semiconductors **35**, 727 (2001)].
5. A. M. Danishevskiĭ, E. L. Ivchenko, S. F. Kochegarov, and V. K. Subashiev, Fiz. Tverd. Tela (Leningrad) **27**, 710 (1985) [Sov. Phys. Solid State **27**, 439 (1985)].
6. P. Schneider, S. D. Ganichev, J. Kainz, *et al.*, Phys. Status Solidi B **238**, 533 (2003).
7. S. D. Ganichev, E. L. Ivchenko, V. V. Bel'kov, *et al.*, Nature **417**, 153 (2002).
8. S. A. Tarasenko, E. L. Ivchenko, V. V. Bel'kov, *et al.*, in *Proceedings of 26th International Conference on Physics of Semiconductors* (Edinburgh, UK, 2002), CD.
9. S. A. Tarasenko, E. L. Ivchenko, V. V. Bel'kov, *et al.*, J. Supercond. **16**, 419 (2003).
10. S. D. Ganichev, V. V. Bel'kov, P. Schneider, *et al.*, Phys. Rev. B **68**, 035319 (2003).
11. S. D. Ganichev, P. Schneider, V. V. Bel'kov, *et al.*, Phys. Rev. B **68**, 081302 (2003).
12. S. D. Ganichev and W. Prettl, J. Phys.: Condens. Matter **15**, R935 (2003).
13. E. L, Ivchenko, Yu. B. Lyanda-Geller, and G. E. Pikus, Zh. Éksp. Teor. Fiz. **98**, 989 (1990) [Sov. Phys. JETP **71**, 550 (1990)].
14. G. E. Pikus, V. A. Marushchak, and A. N. Titkov, Fiz. Tekh. Poluprovodn. (Leningrad) **22**, 185 (1988) [Sov. Phys. Semicond. **22**, 115 (1988)].
15. R. Q. Yang, J. M. Xu, and M. Sweeny, Phys. Rev. B **50**, 7474 (1994).
16. R. J. Warburton, C. Gauer, A. Wixforth, *et al.*, Phys. Rev. B **53**, 7903 (1996).
17. E. E. Takhtamirov and V. A. Volkov, Phys. Low-Dimens. Semicond. Struct. **1**/**2**, 95 (1997).
18. M. A. Zudov, A. P. Mitchell, A. H. Chin, and J. Kono, J. Appl. Phys. **94**, 3271 (2003).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 2, 2004, pp. 386–393.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 2, 2004, pp. 435–443.

SOLIDS
Electronic Properties

The Nature and Mechanism of Charging of Electron Traps in $Lu_2SiO_5:Ce^{3+}$ Crystals

N. V. Znamenskiĭ[a], E. A. Manykin[a], E. A. Petrenko[a], T. G. Yukina[a,*], Yu. V. Malyukin[b,], P. N. Zhmurin[b], B. V. Grinev[b], A. A. Masalov[b], and A. P. Shpak[c]**

[a]*Russian Research Centre Kurchatov Institute, pl. Kurchatova 1, Moscow, 123182 Russia*
[b]*Institute of Single Crystals, National Academy of Sciences of Ukraine, Kharkov, 61001 Ukraine*
[c]*Institute of Metal Physics, National Academy of Sciences of Ukraine, Kiev, 03680 Ukraine*
*e-mail: *zarja@isssph.kiae.ru; **malyukin@isc.kharkov.com*
Received February 17, 2004

Abstract—Investigation of the thermoluminescence (TL) properties depending on the temperature, UV irradiation dose, and activator ion concentration in $Lu_2SiO_5:Ce^{3+}$ (LSO:Ce) crystals and measurements of the decay of recombination radiation during photostimulated release of the excitation energy accumulated in these crystals showed that a part of Ce^{3+} impurity ions exhibit ionization leading to the injection of electrons into the conduction band. The conduction band of a LSO:Ce crystal is involved in two opposite processes: the charging of electron traps and the recombination of electrons with Ce^{4+} ions. The transport of electrons to the traps has a diffusion character: electrons possess a significant mobility [D(315 K) = 10^{-3} cm^2/s] and can diffuse away from the donor ions to distances exceeding the lattice parameter of the crystal. The results of experiments with controlled atomic packing of LSO:Ce nanoclusters unambiguously demonstrated the key role of the structure of oxyorthosilicates in the formation of electron traps. The existing two-center model does not provide adequate description of the properties of electron traps and TL in LSO:Ce crystals. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In recent years, increasing attention has been devoted to the properties of crystals activated by Ce^{3+} ions [1–11]. This interest is related to the search for fast scintillators operating on the *f–d* electron transitions in rare earth element (REE) ions [1–5]. The best scintillation characteristics (the conversion efficiency and the luminescence time) among the materials doped with Ce^{3+} ions were demonstrated by $Lu_2SiO_5:Ce^{3+}$ (LSO:Ce) crystals belonging to oxyorthosilicates of the M_2SiO_5 group (where M is a REE ion) [2, 3]. However, it was also established that γ-irradiation of the LSO:Ce crystals leads to effective accumulation of the energy, which can be subsequently released in the form of afterglow or (on heating the crystal) thermoluminescence (TL) [3, 8, 10, 11]. Both these phenomena significantly restrict the practical applications of LSO:Ce crystals in image display devices [8]. In view of the high potential of oxyorthosilicates activated with Ce^{3+} ions, it is important to elucidate the mechanism of energy accumulation in these materials and construct an adequate physical model explaining the observed phenomena. The existing two-center model, based on the notions about Ce^{4+} ions and localized electrons e^-, was proposed by Dorenbos *et al.* [11]. However, this model has recently met significant objections by Cooke *et al.* [8] and will be the main object of discussion in our paper.

In the two-center model [11], the localization of electrons near excited Ce^{3+} ions was described in terms of a multiwell adiabatic potential reflecting specific features of relaxation processes in the system under consideration: (i) oxygen ions in the vicinity of an excited Ce^{3+} ion may change their equilibrium arrangement and acquire an energetically more favorable configuration with the formation of a potential well and (ii) Ce^{3+} ion can lose one electron ($Ce^{3+} \longrightarrow Ce^{4+} + e^-$) that is trapped by the formed potential well (dynamic trap). The trapped electron forms a metastable state responsible for both afterglow and TL: these processes accompany the reverse process of electron recombination with Ce^{4+} ion [11]. Thus, the essence of the two-center model [11] is that the same cerium ion is the donor of an electron and the center of recombination. It should be emphasized that, according to this model, the excited Ce^{3+} ion is the center of the electron trap formation.

Subsequently [8], it was established that the afterglow and TL are inherent in several other crystals of oxyorthosilicates of the M_2SiO_5 group (M = Lu, Y, Yb, Er, and Gd) doped with Ce^{3+} ions, where the position of the main TL peak was dependent on the type of the particular crystal lattice (M_2SiO_5). An additional, substantially important, but contradictory result was that the pure (undoped) LSO crystals also exhibited afterglow and TL upon γ-irradiation [8]. The discrepancy with the

existing model was that the TL spectrum of LSO free of Ce^{3+} ions coincided with the TL spectrum of a LSO:Ce crystal. Nevertheless, it was ascertained [8] that (i) an excited Ce^{3+} ion is not the center of electron trap formation, (ii) such traps exist in LSO:Ce independently of Ce^{3+} ions, and (iii) the traps can be filled with electrons as a result of relaxation of the γ-irradiation-induced electron–hole pairs, rather than only due to the $Ce^{3+} \longrightarrow Ce^{4+} + e^-$ process. Several configurations involving REE ion surrounded by oxygen atoms suggested as a possible model of the electron trap [8] were rather close to the model considered in [11]. The conclusions derived in [8] were not convincing because neither the nature of electron donors nor the mechanism of electron transport to the traps were considered. Anyhow, the experimental results obtained in [8] cast doubt with respect to adequacy of the two-center model [11].

In this paper, it will be shown that, during optical excitation in the impurity absorption band of an LSO:Ce crystal, a certain fraction of Ce^{3+} activator ions exhibit ionization by the $Ce^{3+} \longrightarrow Ce^{4+} + e^-$ scheme leading to charging of the electron traps. The donated electron can diffuse away from the ion donor to distances exceeding the lattice parameter of the crystal and recombine with another Ce^{4+} ion, which implies that the two-center model developed in [11] is inadequate. The transport of electrons to the traps has a diffusion character and is ensured by the conduction band of the LSO:Ce crystal. The results of experiments with controlled atomic packing of LSO:Ce nanoclusters unambiguously indicate that electron traps are related to the structural features of oxyorthosilicates and confirm that excited Ce^{3+} ion is not the center of electron trap formation.

2. EXPERIMENTAL METHODS

The optical absorption and luminescence spectra of crystals were measured on an automated spectrofluorimeter based on an MDR-23 grating monochromator. The measuring system was controlled by a personal computer using an original software package. The electronic control unit of the monochromator step motor drive, pulse counters, time-to-amplitude converter, and analog-to-digital converter were implemented in the CAMAC standard. The computer was linked to the CAMAC crate via a specially designed controller assuring high speed of the control signal and data transmission.

The stationary luminescence spectra were measured using a FÉU-100 photoelectron multiplier operating in the photon counting mode [12]. The luminescence decay curves were measured in the time-correlated photon counting regime [12] using the scheme based on a FEU-164 photomultiplier. Cerium impurity ions were excited by second harmonic pulses of an organic dye laser synchronously pumped by the second harmonic radiation of a YAG:Nd laser operating in the regime of active synchronization of longitudinal modes.

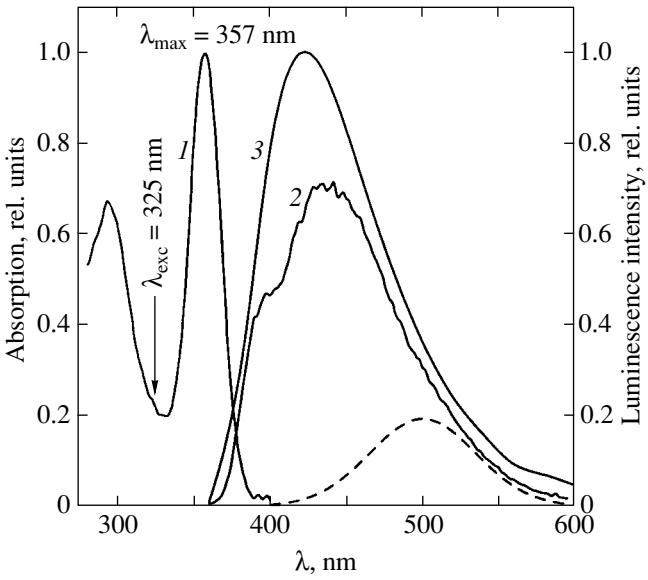


Fig. 1. The spectra of (*1*) impurity absorption, (*2*) luminescence at T = 300 K, and (*3*) TL of an LSO:Ce crystal. Dashed curve shows the luminescence of Ce^{3+} centers of the second type.

Spatially resolved measurements were performed with an original microscope capable of providing the required laser beam focusing and positioning in a sample crystal. LSO:Ce crystals were grown by B.I. Minkov (Institute of Single Crystals, Kharkov). LSO:Ce nanocrystals were obtained using sol-gel technology and examined by electron microscopy.

3. EXPERIMENTAL RESULTS AND DISCUSSION

As is known [13–16], impurity ions can occupy two nonequivalent cationic sites in the crystal lattice of oxyorthosilicates. Accordingly, LSO:Ce crystals contain optical centers of two types with Ce^{3+} ions [2, 9]. The optical absorption spectrum of our LCO:Ce crystal shown in Fig. 1 is identical with the spectra reported in the literature [9]. The optical absorption of Ce^{3+} ions is due to the *f–d* electron transitions [9]. The intensity of absorption for one type of the Ce^{3+} optical centers is significantly higher than that of the other type, and the corresponding spectral bands are not resolved [2, 9]. Excitation at an energy corresponding to the maximum of the long-wavelength absorption band (Fig. 1) leads to predominant excitation of the Ce^{3+} optical centers of the first type and is accompanied by intense emission from these centers with a peak at 420 nm and a relatively weak luminescence from Ce^{3+} optical centers of the second type with a maximum at 500 nm (Fig. 1).

After a certain exposure of the LSO:Ce crystals to radiation within the impurity absorption band (Fig. 1),

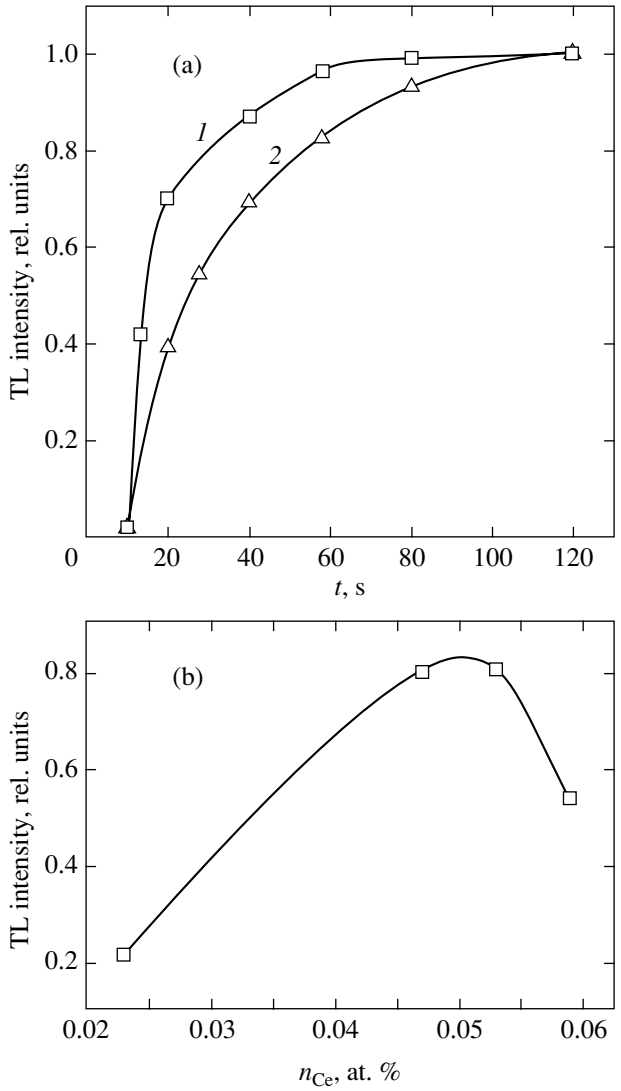


Fig. 2. TL intensity saturation (a) with increasing exposure during optical excitation of LSO:Ce crystals with different Ce^{3+} ion concentrations (*1*, 0.054 at.%; *2*, 0.023 at.%) and (b) with increasing Ce^{3+} activator ion concentration.

it was possible to measure a TL spectrum with the main peak at 375 K, analogous to the spectra observed upon γ-irradiation [3, 11]. In the case of optical excitation, only Ce^{3+} ions play the role of electron donors charging the electron traps responsible for the TL. The optically excited Ce^{3+} ion passes into a higher oxidation state (Ce^{4+}). The identity of thermo- and photoluminescence spectra (Fig. 1) indicates that cerium ions also play the role of recombination centers.

We have established that, under certain conditions, optical excitation of the impurity ions leads to saturation of the TL intensity in LSO:Ce crystals (Fig. 2). In order to provide for a uniform distribution of charged traps in the sample volume, we performed experiments on LSO:Ce crystals with a thickness of about 1 mm. The crystals were irradiated by an unfocused beam of a HeCd laser (λ_{exc} = 325 nm; power, 1 mW; spot diameter, ~1 mm). The position of the laser excitation line is indicated by the arrow in Fig. 1. Prior to irradiation, LSO:Ce crystals were annealed at 500 K until complete vanishing of the TL signal. Then, the samples were cooled to room temperature and exposed to laser radiation for a certain period of time. The TL intensity exhibited saturation with increasing exposure, which was more rapidly reached in LSO:Ce crystals with a greater concentration of activator ions (Fig. 2a).

Since the density of electrons generated in the crystal volume is proportional to the concentration of excited Ce^{3+} ions, it is necessary to analyze variation of the content of such ions and the degree of excitation reached in the ensemble of Ce^{3+} impurity ions. Assuming that the irradiated region is homogeneous in the longitudinal direction, the time evolution of the concentration of Ce^{3+} ions can be described by a simple kinetic equation,

$$\frac{dn^*_{Ce}}{dt} = -\gamma n^*_{Ce} + n_{Ce}\sigma F. \qquad (1)$$

Here, n^*_{Ce} is the concentration of excited Ce^{3+} ions, n_{Ce} is the concentration of Ce^{3+} ions, σ is the optical absorption cross section for the *f–d* electron transition in Ce^{3+} ion, F is the photon flux density per unit time, and γ is a constant coefficient describing the decay of the excited state of Ce^{3+} ions.

For the initial condition $n^*_{Ce}(0) = 0$, Eq. (1) is satisfied by the solution

$$n^*_{Ce}(t) = \frac{n_{Ce}\sigma F}{\gamma}(1 - e^{-\gamma t}),$$

from which it follows that the concentration of excited Ce^{3+} ions very rapidly (on the experimental time scale) reaches a stationary value of $n^*_{Ce} = n_{Ce}\sigma F/\gamma$ (with the constant $\gamma \sim 10^7$ corresponding to the radiative decay of the excited state of Ce^{3+} ions). The level of excitation in the system is maintained on a very low level, since $\sigma F/\gamma \sim 1$ ($\sigma \sim 10^{-17}$ cm^2, $F \sim 10^{15}$ photons/(cm^2 s)). Therefore, the concentration of charged traps n_{tr} in the volume of the crystal varies with time according to the linear law,

$$n_{tr} = \mu n_{Ce}\sigma F t, \qquad (2)$$

where μ is a coefficient taking into account the probability of formation of a charged electron trap.

According to the two-center model [11], the concentration of charged electron traps is proportional to the concentration of activator ions (Ce^{3+}). Then, according to Eq. (2), the ratio n_{tr}/n_{Ce} is independent of the activator ion concentration and, hence, saturation of the TL intensity should be observed after the same time for LSO:Ce crystals with various concentrations of Ce^{3+} ions. The interaction between Ce^{3+} ions can be ignored

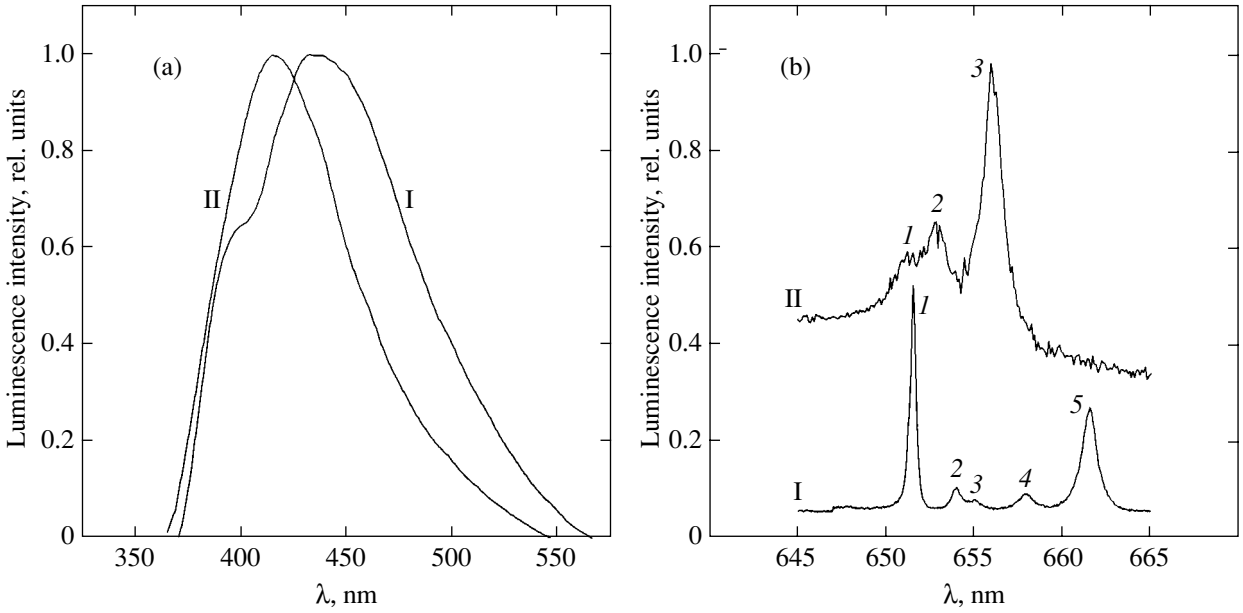


Fig. 3. Luminescence spectra of (I) large and (II) small nanoclusters of (a) LSO:Ce^{3+} and (b) LSO:Pr^{3+}.

since, for 0.023 at. % ≤ n_{Ce} ≤ 0.054 at. %, the distance between neighboring ions is no less than 70 Å. The electron trapped in a potential well is localized near the corresponding impurity ion within a region on the order of 10 Å. Thus, the two-center model cannot describe the dependence of the rate of saturation of the TL intensity on the concentration of activator ions. However, if the electron traps in LSO:Ce crystals exist independently of Ce^{3+} ions, the rate of filling of these traps with electrons must depend on the concentration of these ions (see formula (2)). This dependence is clearly pronounced in Fig. 2a: the higher the concentration of Ce^{3+} ions, the greater the slope of the initial part of the experimental TL intensity curves.

At a constant level of photons absorbed in the crystal volume, a weak saturation of the TL intensity is also observed with increasing concentration of Ce^{3+} ions (Fig. 2b). For an LSO:Ce^{3+} crystal with an activator ion concentration of 0.096 at.%, TL was not observed at all. This fact implies that a distance between Ce^{3+} ions in this case is shorter than the effective radius of electron recombination with Ce^{4+} ion. According to the two-center model [11], the recombination efficiency must be independent of the activator ion concentration.

If the electron traps and Ce^{3+} ions form independent subsystems in an LSO:Ce crystal, the question naturally arises as to how electrons are transported to the traps. At all probability, the distance between an ion donor and the trap is greater than the crystal lattice parameter and, hence, the electron can only surmount this barrier via the conduction band of the LSO:Ce crystal. This is possible only provided that the exited state of Ce^{3+} ions is close to the conduction band bottom and the ionization of this ion is followed by the electron injection into the conduction band. In order to verify this assumption, we attempted to modify in a controlled manner the mutual arrangement of the level of excited Ce^{3+} ions and the conduction band bottom by means of the dimensional quantization (quantum confinement) effect [17]. For this purpose, we have grown LSO:Ce nanocrystals of controlled size. It was found that the characteristics of luminescence of Ce^{3+} ions in large LSO:Ce clusters with dimensions up to 20 nm are identical with the characteristics of LSO:Ce crystals. These large clusters also exhibited TL, which was not observed for small clusters with a size of about 5 nm. After annealing (leading to an increase in the cluster size) the latter clusters also exhibited TL. This behavior was not a direct manifestation of the quantum confinement effect, as confirmed by the following experimental fact. In the small clusters, TL was not observed even after the exposure to short-wavelength photons with λ as small as 200 nm. Some additional experimental facts (see below) also indicated that the mutual arrangement of the bottom of the conduction band in LSO:Ce and the excited level of Ce^{3+} ion is not a decisive factor for the formation of electron traps in LSO:Ce crystals.

Small-size LSO:Ce clusters are simply free of electron traps, which is corroborated by the following experimental data. The impurity luminescence spectrum of small LSO:Ce clusters has a special shape (Fig. 3a), suggesting that such clusters contain Ce^{3+} centers of only one type and, hence, have special atomic packing. Indeed, this was confirmed by probing the cluster structure with Pr^{3+} ions (Fig. 3b). The optical spectra of these ions (in contrast to those of Ce^{3+} ions) are characterized by narrow spectral lines, which can be used as sensitive indicators of the crystal (ligand) field struc-

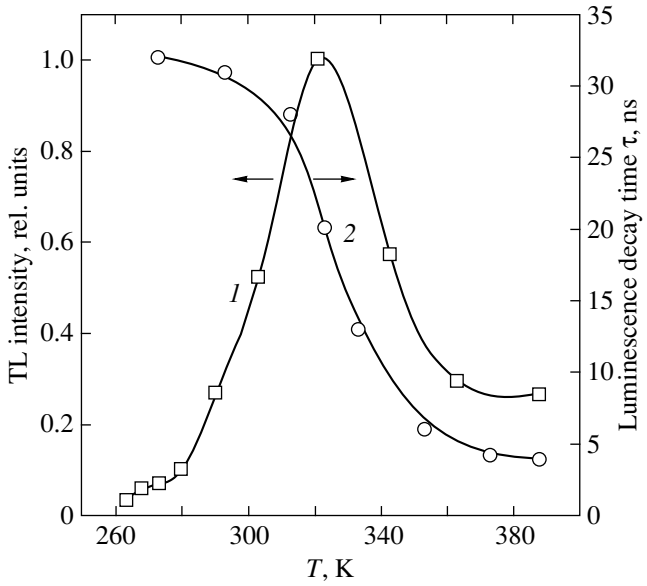


Fig. 4. Temperature dependence of (*1*) the energy accumulation efficiency and (*2*) the luminescence decay time for the excited 5*d* state of Cu^{3+} activator ions in an LSO:Ce crystal.

ture in LSO nanoclusters. The TL spectra of LSO:Pr^{3+} clusters with characteristic dimensions up to 20 nm, as well as the spectra of LSO:Pr^{3+} crystals [13–15], revealed two types of Pr^{3+} centers, whereas the spectra of small LSO:Pr^{3+} clusters with a characteristic size of 5 nm showed the presence of Pr^{3+} centers of only one type.

The impurity luminescence spectrum of small clusters exhibited a clearly pronounced change in the splitting of terms of the Pr^{3+} ion. This is illustrated in Fig. 3b, which shows the fragments corresponding to $^3P_0 \longrightarrow {}^3F_2$ transitions in the fluorescence of Pr^{3+} ions in small and large LSO:Pr^{3+} clusters. The $^3P_0$ term is nondegenerate, while the fivefold-degenerate $^3F_2$ term exhibits splitting in the crystal field and each Stark component can be manifested by a spectral component in the luminescence spectrum. Indeed, because of a low point symmetry of cationic sites in the LSO crystal lattice, the $^3F_2$ term of Pr^{3+} ions is split into five Stark components and the spectrum of $^3P_0 \longrightarrow {}^3F_2$ fluorescence displays five spectral lines (Fig. 3b, peaks *1–5*). Thus, Fig. 3b shows a significant change in splitting of the $^3F_2$ term in small LSO:Pr^{3+} clusters, which is unambiguous evidence of modified atomic packing. Moreover, a decrease in the number of spectral lines indicates that degeneracy of the $^3F_2$ term in small clusters is removed incompletely and, hence, the cationic sites possess a higher point symmetry. We may conclude that the absence of TL for small LSO:Ce^{3+} clusters is also related to a change in their atomic packing, which hinders the formation of electron traps. This is again convincing evidence of a key role of the crystal lattice structure in the formation of electron traps in oxyorthosilicates.

For LSO:Ce^{3+} crystals excited in the impurity absorption band (Fig. 1), the efficiency of energy accumulation exhibited significant changes within a narrow temperature interval. We have measured the TL curves in the samples excited by irradiation to the same dose (Fig. 4). The observed decay of the TL intensity at temperatures above 320 K correlates with a decrease in the quantum yield of luminescence of Ce^{3+} ions and in the lifetime of their 5*d* state (Fig. 4). Therefore, the integral TL intensity decreases because of the luminescence quenching for Ce^{3+} ions. At temperatures below 260 K, metastable states are not formed (Fig. 4). However, using radiation of a shorter wavelength (λ_{exc} = 265 nm), such states can be created at temperatures as low as 80 K. Therefore, for Ce^{3+} ions excited in the long-wavelength impurity absorption band (Fig. 1), the mechanism of trap charging has an activation character.

The region of increase in the TL intensity (Fig. 4) is described by exponential function $\exp(\Delta/kT)$ with $\Delta \approx 0.5$ eV. The temperature dependence of the efficiency of energy accumulation in LSO:Ce crystals agrees with the two-center model [11]. The activation energy $\Delta \approx 0.5$ eV can be interpreted as the height of a potential barrier separating two states in a multiwell adiabatic potential. According to an alternative interpretation, this activation energy is equal to the energy between the bottom of the conduction band and the level of excited state of the Ce^{3+} activator ion. However, both these assumptions require additional experiments for determining the mobility of electrons appearing as a result of the photoionization of Ce^{3+} ions.

In order to evaluate the electron mobility, we used a focused beam of laser radiation creating charged electron traps within a limited volume of the LSO:Ce crystal. These traps were detected by measuring the photostimulated release of the accumulated excitation energy [10]. Using probing photon energies ranging from the UV to IR spectral regions, it was possible to empty the traps and observed radiation due to the recombination of released electrons with Ce^{4+} ions. The extremely wide energy spectrum of probing photons also confirms the above assumption concerning a key role of the crystal conduction band in the electron transport in LSO:Ce crystals.

In our experiments, the energy accumulated in an LSO:Ce crystal was released under the action of a He–Ne laser (λ_{exc} = 628 nm; power, 1 mW) incapable of producing charged traps. The charged traps were created by preliminary local irradiation of the crystal by a focused beam of a single-mode He–Cd laser (λ_{exc} = 325 nm; power, 1 mW; spot diameter, ~20 μm). Both laser beams (focused in the same region of the sample) were initially blocked for a period of time required to anneal the crystal until complete vanishing of the TL

signal. Then, the He–Cd laser beam was unblocked and the crystal was activated by 325-nm radiation for a certain time at room temperature. Finally, the He–Cd laser beam was blocked and the probing He–Ne laser beam was unblocked to induce emission as a result of the $e^- + Ce^{4+} \longrightarrow Ce^{3+}$ recombination process. The intensity of this emission decreased with time, the character of this decay being dependent on the character of laser beam focusing. If the probing He–Ne laser beam diameter was greater than the He–Cd laser beam spot size, the emission intensity must decay to zero according to the law $\exp(-\delta F t)$, where δ is the absorption cross section of electron traps. The results of such measurements confirmed the dependence of the rate of luminescence intensity decay on the laser radiation power density. If the probing He–Ne laser beam diameter (~10 μm) was smaller than the size of the He–Cd laser beam spot, the curve of the photostimulated emission intensity practically exhibited a plateau and did not fall to zero for a rather long time.

The decay of the photostimulated emission intensity was also affected by preliminary heat treatment of the crystals. This is illustrated by the data in Fig. 5, where each curve is satisfactorily described by the sum of two exponents (see the legend to figure). For the LSO:Ce crystal annealed in an oxygen-containing atmosphere (and, hence, enriched with Ce^{4+} ions), the initial intensity of emission was (under otherwise equal conditions) lower and the decay rate was higher (Fig. 5, curve *2*). This behavior is explained by the fact that electrons released from the initially charged traps more rapidly find a recombination center (Ce^{4+} ion) in the annealed crystal. The presence of a plateau in the emission decay curves indicates that electron traps in the region of focusing of the He–Ne laser beam are continuously supplied by electrons from the surrounding region not illuminated by the probing beam.

The process of electron diffusion is most convincingly demonstrated by the following experiment. The beams of the exciting (He–Cd) and probing (He–Ne) lasers were focused on the crystal surface at the points situated at a distance of 400 μm from one another. Initially, both laser beams were blocked and the crystal was annealed until complete vanishing of the TL signal. Then, the crystal was activated at the first point by constant focused 325-nm radiation so as to completely exclude (using special diaphragms) irradiation of the region of focusing of the probing He–Ne laser beam. Finally, the gated He–Ne laser beam probed the crystal at the second point. It is obvious that, from the standpoint of the model assuming electron localization in the vicinity of Ce^{3+} centers, no photostimulated emission can be observed at a point remote from the region enriched by the charged electron traps. However, the intensity of photostimulated emission from the point probed by the He–Ne laser gradually increases (Fig. 6) and the rate of this growth increases with the sample temperature.

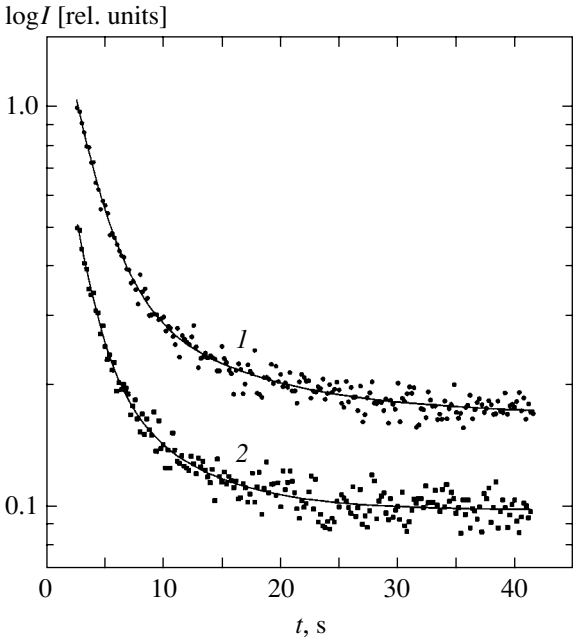


Fig. 5. Decay of the photostimulated luminescence intensity I measured at $T = 300$ K in LSO:Ce crystals with different thermal prehistories (points) and approximation of the experimental data by the law $a_1\exp(-t/\tau_1) + a_2\exp(-t/\tau_2)$: (*1*) unannealed sample, $\tau_1 = 2.5$ s, $\tau_2 = 9.3$ s; (*2*) sample annealed in an oxygen-containing atmosphere, $\tau_1 = 1.5$ s, $\tau_2 = 9.1$ s.

The results of experiments illustrated in Fig. 6 indicate that electrons are transferred from the region of accumulation of the charged traps to the traps significantly remote from the region of injection of free electrons. In the absence of interaction between Ce^{3+} ions, such transport can be explained only by recourse to the band structure of an LSO:Ce crystal. The necessary electron mobility can be provided by the conduction band of the LSO:Ce^{3+} crystal. Judging by the position of the main TL peak (375 K), we may conclude that the electron traps are shallow and they are continuously depopulated as a result of thermal activation. Electrons occurring in the conduction band can be transferred over considerable distances. This explains the recombination radiation observed in the region of illumination by a He–Ne laser, which is not excited by He–Cd laser radiation (Fig. 6).

Using the experimental data presented in Fig. 6, it is possible to estimate the coefficient of diffusion for electrons in LSO:Ce crystals. Let us assume that the electron transport is described by the diffusion equation [18]

$$\frac{\partial n_e}{\partial t} - D\Delta n_e = \rho(x, t), \qquad (3)$$

where n_e is the density of trapped electrons and $\rho(x, t) = \rho\delta(x)$ is the rate of charging of the electron traps (according to Eq. (1), $\rho = \mu n_{Ce}\sigma F$. For the initial condi-

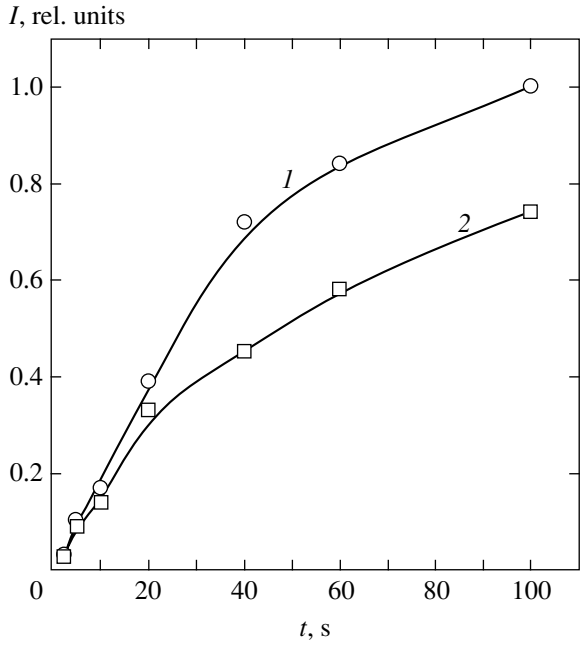


Fig. 6. Buildup of the photostimulated luminescence intensity I measured at T = 313 (*1*) and 288 K (*2*) at a point spaced by 400 μm from the point of charged electron trap formation in an LSO:Ce crystal.

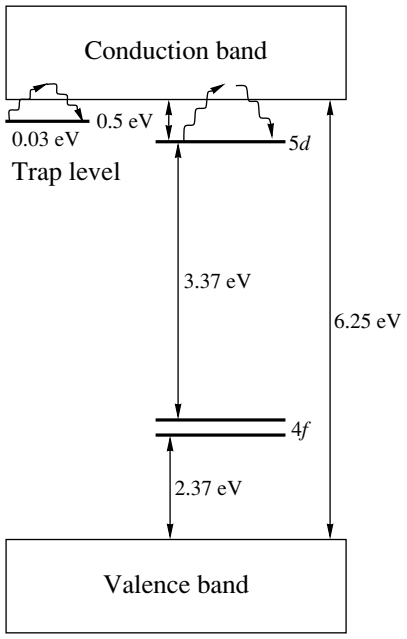


Fig. 7. Simplified energy band diagram showing the mutual arrangement of the ground 4*f* and excited 5*d* states of activator ions in LSO:Ce crystals.

tion $n_e(x, 0) \equiv 0$, a solution of Eq. (3) can be presented in the following form [18]:

$$n_e(x, t) = \int_0^t \rho \frac{\exp[-|x|^2/4D(t-\tau)]}{4D\pi(t-\tau)} d\tau. \quad (4)$$

By substituting the new variable $|x|^2/4D(t-\tau) = \xi$ in the integral in expression (4), this solution can be rewritten as

$$n_e(x, t) = \frac{\rho}{4D\pi} \int_{|x|^2/4Dt}^{\infty} \frac{\exp(-\xi)}{\xi} d\xi. \quad (5)$$

This form is convenient for approximation of the experimental curves presented in Fig. 6. By varying the parameter D in expression (5) for $|x|^2$ = 400 μm^2, we obtain the best approximation of these curves for the following values of the diffusion coefficient:

$$D = 6 \times 10^{-4} \text{ cm}^2 \text{ s}^{-1} \ (T = 280 \text{ K}); \quad (5.1)$$

$$D = 10^{-3} \text{ cm}^2 \text{ s}^{-1} \ (T = 315 \text{ K}). \quad (5.2)$$

In the final interpretation, the activation energy $\Delta \approx$ 0.5 eV should be identified with the energy gap between the level of the excited state of Ce^{3+} ions and the bottom of the conduction band of LSO:Ce. Based on this assumption, we may reconstruct a scheme of the mutual arrangement of energy levels in the ground and excited states of Ce^{3+} ions in LSO:Ce crystals (Fig. 7). The bandgap width (dielectric gap) of LSO:Ce estimated from the position of the UV absorption edge is about 6.25 eV. Within the framework of the model of energy states of the LSO:Ce crystal and impurity ions depicted in Fig. 7, it is possible to provide adequate description of all the experimental results presented above.

4. CONCLUSIONS

During the photoexcitation of LSO:Ce crystals in the impurity absorption band, a part of Ce^{3+} activator ions exhibit ionization leading to the injection of electrons into the conduction band. The activator ions play the role of electron donors and recombination centers, but are not the centers responsible for the formation of electron traps. The two-center model developed in [11] is fully rejected. The formation of charged metastable states and the recombination of Ce^{4+} ions with electrons are the opposite processes proceeding with participation of the conduction band of LSO:Ce crystals. The transport of electrons to the traps has a diffusion character. Electrons can diffuse away from the ion donors to distances exceeding the lattice parameter of the LSO:Ce crystal.

REFERENCES

1. C. L. Melcher, R. A. Manente, C. A. Peterson, *et al.*, J. Cryst. Growth **128**, 1001 (1993).

2. H. Suzuki, T. A. Tombrello, C. L. Melcher, *et al.*, Nucl. Instrum. Methods Phys. Res. A **320**, 263 (1992).
3. P. Dorenbos, C. W. E. van Eijk, A. J. J. Bos, *et al.*, J. Lumin. **60**–**61**, 979 (1994).
4. R. H. Bartram, D. S. Hamilton, L. A. Kappers, *et al.*, J. Lumin. **75**, 183 (1997).
5. J. C. Spijker, P. Dorenbos, C. W. E. van Eijk, *et al.*, J. Lumin. **85**, 1 (1999).
6. B. S. K. Nair, D. Sundar, A. Tomita, *et al.*, J. Lumin. **86**, 67 (2000).
7. U. Happek, J. Choi, and A. M. Srivastava, J. Lumin. **94**–**95**, 7 (2001).
8. D. W. Cooke, B. L. Bennett, K. J. McClellan, *et al.*, J. Lumin. **92**, 83 (2001).
9. H. Suzuki, T. A. Tombrello, C. L. Melcher, *et al.*, IEEE Trans. Nucl. Sci. **40**, 380 (1993).
10. R. Visser, C. L. Melcher, J. S. Schweizer, *et al.*, IEEE Trans. Nucl. Sci. **41**, 689 (1994).
11. P. Dorenbos, C. W. E. van Eijk, A. J. J. Bos, *et al.*, J. Phys.: Condens. Matter **6**, 4167 (1994).
12. D. V. O'Connor and D. Phillipe, *Time-Correlated Single Photon Counting* (Academic, New York, 1984).
13. R. S. Borisov, N. V. Znamenskiĭ, Yu. V. Malyukin, *et al.*, Fiz. Nizk. Temp. **24**, 571 (1998) [Low Temp. Phys. **24**, 432 (1998)].
14. R. S. Borisov, N. V. Znamenskiĭ, Yu. V. Malyukin, *et al.*, Fiz. Nizk. Temp. **26**, 1207 (2000) [Low Temp. Phys. **26**, 894 (2000)].
15. R. S. Borisov, N. V. Znamenskiĭ, Yu. V. Malyukin, *et al.*, Zh. Éksp. Teor. Fiz. **120**, 420 (2001) [JETP **93**, 372 (2001)].
16. R. S. Borisov, B. V. Grinev, Yu. V. Malyukin, *et al.*, Low Temp. Phys. **27**, 574 (2001).
17. U. Kreibig and M. Vollmer, *Optical Properties of Metal Cluster* (Springer, Berlin, 1995).
18. V. S. Vladimirov, *Equations of Mathematical Physics*, 3rd ed. (Nauka, Moscow, 1981; Marcel Dekker, New York, 1971).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 2, 2004, pp. 394–414.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 2, 2004, pp. 444–468.

SOLIDS
Electronic Properties

Spin-Polaron Transport and Magnetic Phase Diagram of Iron Monosilicide

V. V. Glushkov[a,b,*], I. B. Voskoboĭnikov[a], S. V. Demishev[a,b], I. V. Krivitskiĭ[a], A. Menovsky[c], V. V. Moshchalkov[d], N. A. Samarin[a], and N. E. Sluchanko[a,b]

[a] *Prokhorov Institute of General Physics, Russian Academy of Sciences, Moscow, 119991 Russia*
[b] *Moscow Institute of Physics and Technology, Dolgoprudnyĭ, Moscow oblast, 141700 Russia*
[c] *Van der Waals–Zeeman Laboratory, University of Amsterdam, 1018 XE, Amsterdam, the Netherlands*
[d] *Laboratorium voor Vaste-Stoffysica en Magnetisme, Katholieke Universiteit Leuven, B-3001, Leuven, Belgium*
**e-mail: glushkov@lt.gpi.ru*
Received February 17, 2004

Abstract—The study of galvanomagnetic, magnetic, and magnetooptical characteristics of iron monosilicide in a wide range of temperatures (1.8–40 K) and magnetic fields (up to 120 kOe) has revealed the origin of the low-temperature sign reversal of the Hall coefficient in FeSi. It is shown that this effect is associated with an increase in the amplitude of the anomalous component of the Hall resistance ρ_H (the amplitude increases by more than five orders of magnitude with decreasing temperature in the range 1.8–20 K). The emergence of the anomalous contribution to ρ_H is attributed to the transition from the spin-polaron to coherent regime of electron density fluctuations in the vicinity of Fe centers and to the formation of nanosize ferromagnetic regions, i.e., ferrons (about 10 Å in diameter), in the FeSi matrix at $T < T_C = 15$ K. An additional contribution to the Hall effect, which is observed near the temperature of sign reversal of ρ_H and is manifested as the second harmonic in the angular dependences $\rho_H(\varphi)$, cannot be explained in the framework of traditional phenomenological models. Analysis of magnetoresistance of FeSi in the spin-polaron and coherent spin fluctuation modes shows that the sign reversal of the ratio $\Delta\rho(H)/\rho$ accompanied by a transition from a positive ($\Delta\rho/\rho > 0$, $T > T_m$) to a negative ($\Delta\rho/\rho < 0$, $T < T_m$) magnetoresistance is observed in the immediate vicinity of the mictomagnetic phase boundary at $T_m = 7$ K. The linear asymptotic form of the negative magnetoresistance $\Delta\rho/\rho \propto -H$ in weak magnetic fields up to 10 kOe is explained by the formation of magnetic nanoclusters from interacting ferrons in the mictomagnetic phase of FeSi at $T < T_m$. The results are used for constructing for the first time the low-temperature magnetic phase diagram of FeSi. The effects of exchange enhancement are estimated quantitatively and the effective parameters characterizing the electron subsystem in the paramagnetic ($T > T_C$), ferromagnetic ($T_m < T < T_C$), and mictomagnetic ($T < T_m$) phases are determined. Analysis of anomalies in the aggregate of transport, magnetic, and magnetooptical characteristics observed in the vicinity of $H_m \approx 35$ kOe at $T < T_m$ leads to the conclusion that a new collinear magnetic phase with **M** || **H** exists on the low-temperature phase diagram of iron monosilicide. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Peculiar electrophysical and magnetic properties of the narrowband FeSi semiconductor determine its special place among the objects studied by the physics of magnetic phenomena. Iron monosilicide is one of the most striking examples of the successful application of the self-consistent renormalization theory of spin fluctuations [1, 2] explaining the effect of formation and enhancement of temperature-induced localized magnetic moments at Fe centers in the FeSi matrix at temperatures above 100 K [1–4]. The results of band analysis [5–8] and experiments carried out in ultrastrong magnetic fields [9, 10] indicate that FeSi is also a convenient model object for studying field-induced metamagnetism in a system of collectivized electrons. In accordance with the conclusions drawn in [11, 12], iron monosilicide and solid solutions on its basis are of considerable interest for studying the metal–insulator transitions in a system with heavy fermions. Finally, in spite of the half-century history of experiments with this compound with a simple cubic crystallographic structure (of the *B*20 type) (see, for example, [13]), the nature and the features of the formation of the ground state and appropriate approaches to a theoretical description of the low-temperature properties of FeSi have been actively discussed in recent years [14–21].

It was shown on the basis of comparatively recent results of measurements of magnetic, transport, and thermoelectric characteristics [22, 23] that the most adequate approach to the description of the low-temperature ground state is apparently interpretation in terms of the Mott–Hubbard model [24]. In this

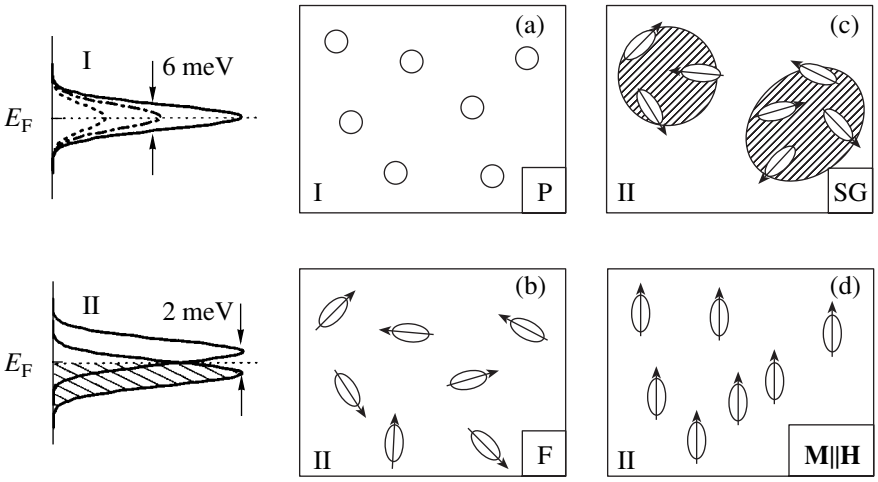


Fig. 1. Various types of many-particle states in the FeSi matrix: (a) the paramagnetic phase with spin polarons as charge carriers (phase P, 15 K < T < 100 K); (b) the phase of noninteracting ferromagnetic nanosize regions (phase F, 7 K < T < 15 K); (c) the mictomagnetic state (spin glass phase) (SG, $T \le 7$ K, $H < 35$ kOe); and (d) the phase with magnetic moments oriented along the external magnetic field (**M** || **H**, 35 kOe < $H \le 350$ kOe). The schematic diagrams on the left show the structure of a many-particle resonance in the immediate vicinity of the Fermi level E_F in the absence (structure I, panel (a)) and in the presence (structure II, panels (b), (c), and (d)) exchange splitting.

approach, at $T < 100$ K, FeSi should be treated as a metal with strong electron correlations and with the spin-polaron regime of charge carrier transport [22, 23]. In particular, the authors of this paper showed [23] that many-particle (spin-polaron) states with a considerably larger effective mass $m^* \sim 100m_0$ (m_0 is the free electron mass) are formed around charge carriers in the upper Hubbard (t_{2g}) band of the narrow-band semiconductor FeSi as its temperature decreases in the interval $T < 100$ K. Such heavy fermions formed as a result of fast electron density fluctuations determine to a considerable extent the low-temperature transport and thermodynamic parameters of FeSi. Further cooling in the system of spin polarons leads to a transition to the coherent regime of spin fluctuations accompanied by the formation of nanosize anisotropic ferromagnetic regions (ferrons) at $T_C \approx 15$ K (T_C is the Curie temperature of the ferromagnetic "phase transition" in the system of spin polarons) and then initiates the formation of a mictomagnetic ground state at $T < T_m \approx 7$ K (T_m is the temperature of the transition of the ferron system to the spin glass phase) [23]. This sequence of transformations of many-particle states in the FeSi matrix is shown schematically in Fig. 1, which depicts the low-temperature phases of iron monosilicide (noninteracting spin polarons are denoted by circles in Fig. 1a; ferrons with uniaxial magnetic anisotropy are shown by ellipses with the easy magnetization axis in Fig. 1b; clusters of interacting ferrons are hatched in Fig. 1c).

In such a situation, the characteristics of the ground state of FeSi are directly connected with the peculiarities of renormalization of the band structure (formation of the many-particle spin-polaron resonance) in the vicinity of the Fermi energy E_F as well as with a complex rearrangement of the magnetic system in the regime of a low density of charge carriers (10^{17}–10^{18} cm^{-3}). The uniqueness of iron monosilicide is associated with the fact that the measurement of galvanomagnetic properties of FeSi in the spin-polaron regime of the low-temperature transport is the most sensitive method for studying the features of stabilization of the coherent regime of spin fluctuations and transformation of many-particle states in nanosize ferromagnetic regions and nanoclusters [23, 25]. At the same time, reliable and authentic experimental data on the behavior of the Hall coefficient $R_H(H, T)$ in this narrowband FeSi semiconductor have not been obtained until recently. As was noted earlier [25], previous measurements of $R_H(H, T)$ pertained to different temperature intervals ($T \ge 4.2$ K [12], 0.05 K $\le T \le$ 55 K [15], and $T \ge 20$ K [26]) and did not match even in the sign of the Hall effect. In addition, the results obtained on the behavior of the field dependences of the Hall resistance were contradictory [12, 15, 26] and the approaches to the interpretation of experimental results were noticeably different. In [25], it was shown for the first time that a decrease in temperature leads to a change in the regimes of charge transport, which in turn leads to double sign inversion in the Hall coefficient of FeSi at $T_{inv1}^H \approx 75$ K and $T_{inv2}^H \approx 12$–15 K. The results of recent experiments [23] enabled us to attribute these anomalies to the transition from the regime of intrinsic conductivity to transport over spin-polaron states in the gap and, further, to the transport of carriers under the conditions of formation of a complex magnetic structure in the FeSi matrix containing ferromagnetic nanoclusters. Considerable advances have been made due to the application of the method of rotation with step-by-step

fixation of the position of the sample in a magnetic field for studying the Hall effect in FeSi [23]. This makes it possible to single out the anomalous (magnetic) component in the Hall resistance and to separate the contributions from two groups of charge carriers to the Hall coefficient at helium ($T \le 4.2$ K) temperatures. In addition, using this approach for the temperature interval corresponding to the spin-polaron transport ($T <$ 100 K), we estimated the value of the exchange field at charge carriers in the upper Hubbard band as $H_{ex} = 350 \pm 100$ kOe [23]. At the same time, the behavior of the field and temperature dependences of the Hall coefficient in FeSi in the region of intermediate temperatures of 4.2–30 K, where, according to Pashen *et al.* [15], "the Hall resistance changes chaotically," could not been investigated properly in view of methodical limitations. In particular, the main difficulties in organization of magnetic and galvanomagnetic cryogenic experiments were associated with the insufficient accuracy (of about 0.05 K) in stabilization and control of temperature in the volume with the sample, which did not allow the authors [23] to study in detail the processes of charge transfer and to determine the structure of the magnetic phase diagram of FeSi.

In the present study, precision measurements of galvanomagnetic and magnetic properties of FeSi single crystals were studied in a wide range of temperatures (1.6–40 K) and magnetic fields (up to 120 kOe) to determine the nature of magnetic interactions and features of the *H–T* magnetic phase diagram of FeSi (including the genesis of the anomalous Hall effect); the anomalies in the microwave magnetoabsorption are also analyzed in the vicinity of low-temperatures magnetic phase transitions.

2. EXPERIMENTAL TECHNIQUE

The measurements of transport, magnetic, and magnetooptical characteristics reported here were made on FeSi single crystals grown from melt in accordance with the Czochralski method and used in our previous experiments [22, 23, 25].

The resistivity and the Hall coefficient were measured on an original experimental setup whose block diagram is described in detail in [27]. The angular dependences of galvanomagnetic characteristics were measured using a scheme with step-by-step rotation of the sample in a magnetic field controlled by a drive from a step motor and with a rotation discreteness of 1.8°–3.6°. After the rotation of the holder with the sample through a fixed angle, signals were measured from the Hall contacts using the standard four-probe scheme. The resistance and magnetoresistance of FeSi samples were also measured using the standard dc four-probe scheme, the behavior of the transverse magnetoresistance being measured in a magnetic field perpendicular to the direction of the current in the sample. Precision measurements of the above resistive characteristics of FeSi samples were carried out using a two-channel nanovoltmeter Keithley 2182.

Experiments in a magnetic field in a wide temperature range (1.6–40 K) were carried out in a vacuum-seal ampoule placed in the channel of a superconducting magnet in a helium cryostat. The temperature in the measuring cell containing the sample was stabilized (to within ~0.01 K) using a digital temperature controller of original construction, which was developed and constructed at the Institute of General Physics, Russian Academy of Sciences. The standard resistance thermometer CERNOX 1050 used as a temperature sensor ensured the required precision of measurements (better than 0.01 K) and minimized the thermometric error associated with a change in the sensor resistance in a magnetic field up to 70 kOe. Data collection and processing and the control of parameters and working regimes of devices and blocks constituting the experimental setup were carried out by the recording and control system based on a microprocessor interfaced with a PC. The magnetoresistance measurements in a magnetic field up to 400 kOe were made on a pulsed-field setup at the Laboratory of Solid-State Physics and Magnetism, Catholic University of Leuven.

The magnetization of FeSi samples was measured on a vibrating-coil magnetometer Oxford Instruments VSM12/V in a temperature range of 1.5–300 K in a magnetic field up to 120 kOe.

The analysis of the microwave magnetoabsorption at low temperatures ($T < 6$ K) in magnetic fields up to 70 kOe was aimed at obtaining additional information on the structure of the magnetic phase *H–T* diagrams of FeSi was carried out using an automated experimental setup of the original design, described in [28]. This technique was successfully used earlier for studying magnetic phase transitions in compounds with strong electron correlations (magnetic Kondo lattices of $CeAl_2$ and CeB_6 [28, 29] as well as the intermediate-valence superconductor $CeRu_2$ [30]). In the present study, the microwave absorption was recorded using a bridge bolometric circuit with a direct contact between the FeSi sample and one of low-temperature bolometers.

3. EXPERIMENTAL RESULTS

3.1. Magnetoresistance

The field dependences of the magnetoresistance of FeSi measured in a magnetic field up to 70 kOe are shown in Figs. 2 and 3. At $T \ge 30$ K, FeSi is characterized by a positive magnetoresistance, which can be described to a high degree of accuracy by a quadratic functions of the form $\Delta\rho/\rho \approx AH^2$ (Fig. 2) in magnetic fields $H \le 70$ kOe used in our experiments. With decreasing temperature, the amplitude of the positive magnetoresistance of FeSi increases and attains its maximal value $\Delta\rho/\rho \approx 3\%$ in a magnetic field of 70 kOe in the vicinity of temperature $T = T_{max}^{\Delta\rho} \approx 11$ K (see

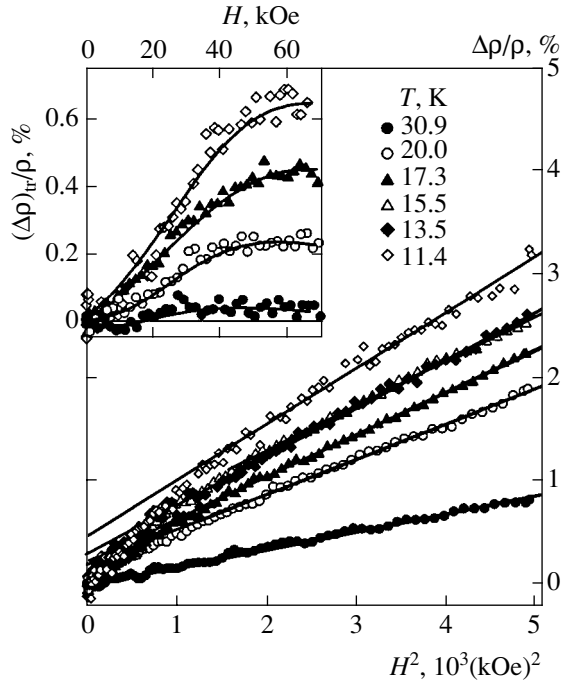


Fig. 2. Field dependences of the magnetoresistance of FeSi in the temperature range 11 K < T < 40 K. The linear approximation corresponds to the quadratic dependence of the magnetoresistance $\Delta\rho/\rho \propto H^2$. The inset shows the family of curves $(\Delta\rho/\rho)_{tr}$ determining the additional positive contribution to the magnetoresistance (see text).

Fig. 2). Further cooling suppresses the positive magnetoresistance and leads to the emergence of a negative contribution to the magnetoresistance at temperatures $T \le T_m \approx 7$ K (Fig. 3). The observed change in the behavior of the magnetoresistance in FeSi (transition from positive to negative magnetoresistance) is graphically illustrated in Fig. 4 showing the temperature dependences of the amplitudes of the positive and negative contributions to the magnetoresistance of FeSi.

Analyzing the features of the magnetoresistive effect in FeSi, we note that a decrease in temperature from 20 to 11 K is not only accompanied by an increase in the amplitude of the quadratic magnetoresistance component, but also leads to noticeable deviations from the quadratic law (see Fig. 2). Subtracting from the experimental curves $\Delta\rho/\rho = f(H)$ of contributions $\Delta\rho \propto H^2$ (solid lines in Fig. 2) makes it possible to single out the additional positive contribution to the magnetoresistance $(\Delta\rho/\rho)_{tr}$ in the range of fields $H \le 35$ kOe. In this case, the family of curves $(\Delta\rho/\rho)_{tr} = \Delta\rho/\rho - AH^2$ (see the inset to Fig. 2) exhibits a structure in the form of a step with an amplitude attaining its maximal value in the vicinity of $T = T_{max}^{\Delta\rho} \approx 11$ K as in the case of the quadratic contribution (curve *2* in Fig. 4). It should also be noted that the steepest segment on the curve describing the increase in the value of $(\Delta\rho/\rho)_{tr}$ in a magnetic

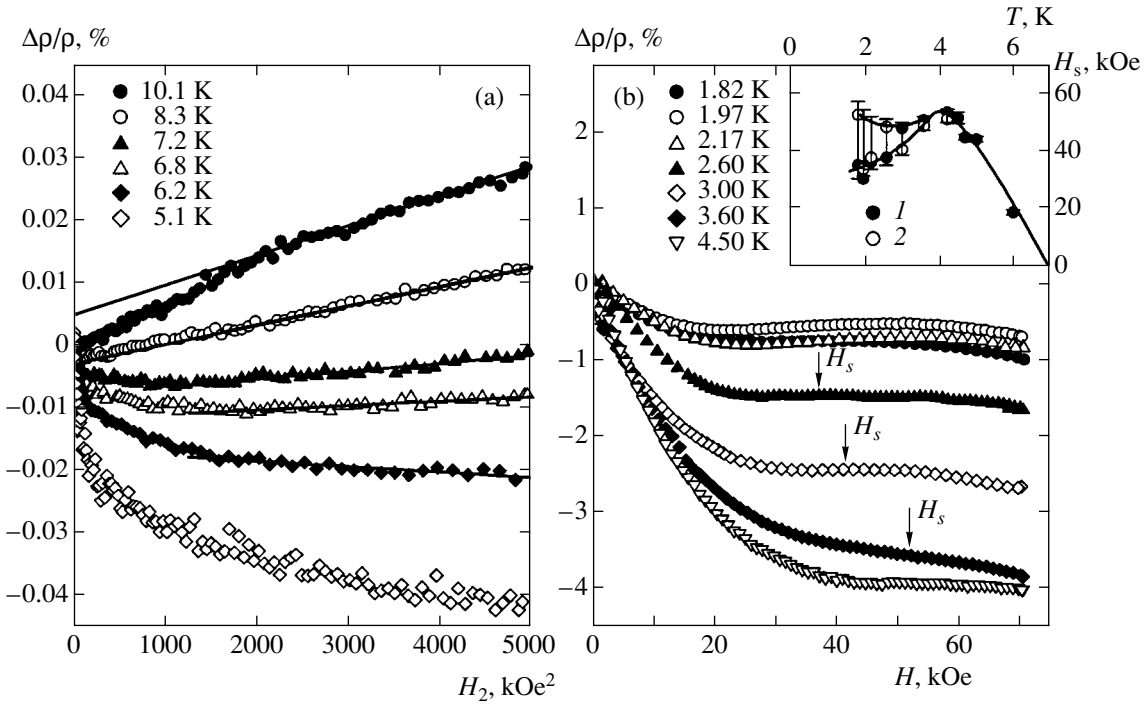


Fig. 3. (a) Field dependences of the magnetoresistance of FeSi in the temperature range 5 K < T < 11 K (straight lines correspond to the quadratic contribution of magnetoresistance $\Delta\rho/\rho \propto H^2$). (b) Field dependences of the negative magnetoresistance of FeSi at T < 5 K. The inset shows the temperature dependences of the magnetic field H_s corresponding to the point of inflection on the curve $\Delta\rho/\rho = f(H)$, obtained in experiments in an increasing (*1*) and decreasing (*2*) magnetic field.

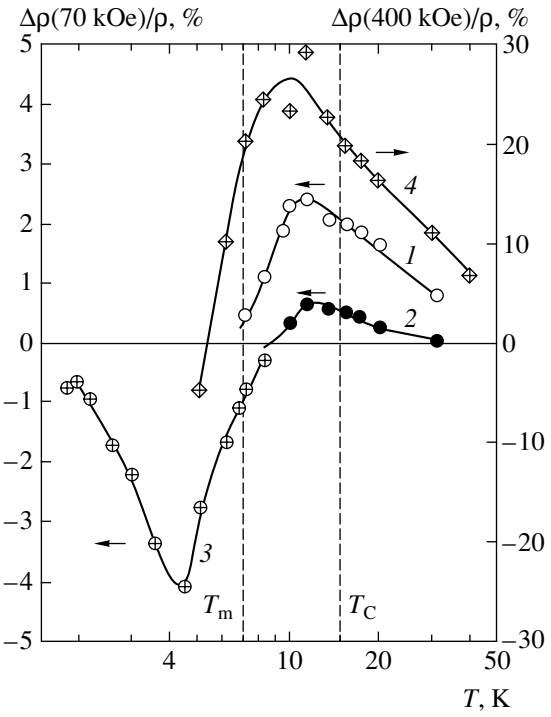


Fig. 4. Temperature dependences of the quadratic contribution AH^2 to the magnetoresistance (*1*), of the nonlinear component $\Delta\rho/\rho - AH^2$ (*2*), the amplitude of the negative contribution A_0 to the magnetoresistance in a magnetic field H = 60 kOe (*3*), and of the magnetoresistance in a pulsed field H = 400 kOe (*4*).

field corresponds to the interval 15–35 kOe (see the inset to Fig. 2).

A decrease in the temperature in the interval of $T \leq 11$ K leads to simultaneous suppression of both above contributions and subsequently to the magnetoresistance sign inversion (see Figs. 3 and 4) in the vicinity of the mictomagnetic transition temperature $T_m \approx 7$ K in the system of ferromagnetic nanoclusters in the FeSi matrix [23]. It can be seen from Fig. 3b that the strongest increase in the amplitude of the negative magnetoresistance is observed in magnetic fields $H \leq 35$ kOe followed by a transition to a regime with a slower variation of $\Delta\rho/\rho$ in the field interval 35–70 kOe.

It is interesting to note that the negative magnetoresistance of FeSi in weak fields ($H \approx 10$ kOe) at temperatures $T < T_m \approx 7$ K turns out to be a linear function of the magnetic field: $\Delta\rho/\rho \propto H$ (see Fig. 3b). The maximum absolute value of the negative magnetoresistance is observed at temperatures of 4–5 K (curve *3* in Fig. 4); as the temperature decreases below the helium temperature, the amplitude of the negative magnetoresistance decreases substantially. It should also be noted that the field dependences of the negative magnetoresistance in FeSi at $T \leq 7$ K in the field range under investigation ($H \leq 70$ kOe) are nonmonotonic; in the intermediate range of fields, the experimental curves $\Delta\rho/\rho = f(H)$ acquire a point of inflection H_s (Fig. 3b). The $H_s(T)$ dependence is shown in the inset to Fig. 3b.

3.2. Hall Coefficient

As was noted in the previous section, a detailed analysis of the Hall effect in FeSi was carried out using the method of rotating a sample with step-by-stem fixation of its position in a magnetic field (see the diagram in Fig. 5a). The angular dependences of the Hall resistance $\rho_H(\varphi)$ obtained using this experimental procedure for FeSi in a magnetic field of $H \approx 59.3$ kOe for several fixed values of temperature in the interval 1.6–40 K are shown in Fig. 5a. For better visualization, the curves are displaced relative to one another by a constant value (the factor on the left of each curve makes it possible to compare the amplitudes of the effect). Figure 6 shows families of experimental curves $\rho_H(\varphi)$ for FeSi, measured at several fixed values of magnetic field $H <$ 75 kOe for a number of temperatures in the intervals $T > T_C$ (T = 20.3 K, Fig. 6a); $T_m < T < T_C$ (T = 13.5 K, Fig. 6b), $T < T_m$ (T = 2.5 K, Fig. 6c) and for a temperature in the immediate vicinity of the mictomagnetic transition (T = 8.3 K $\approx T_m$, Fig. 6d).

In the standard situation, the rotation of the sample in a fixed magnetic field H follows the variation of the projection $H_\perp$ of the external magnetic field vector onto the normal to the sample surface (see the diagram in Fig. 5a) in accordance with the harmonic law $H_\perp = H_0\cos\varphi$ and leads to a cosinusoidal dependence of the Hall resistance of the type $\rho_H \propto R_H(T, H)H\cos\varphi$ at a fixed temperature T. However, such a shape of the $\rho_H(\varphi, H_0, T_0)$ curves was observed in the Hall measurements in FeSi only in a limited range of temperatures and magnetic fields. In particular, at temperatures $T \leq$ 5 K, a nearly cosinusoidal dependence of the Hall signal is observed in fields above 50 kOe (see the curves for T = 2.5 K and T = 5.1 K in Fig. 5a). It should also be noted that a Hall signal of the form $\rho_H \propto \cos\varphi$ is detected only for temperatures above 35 K, when the shape of the $\rho_H(\varphi)$ curves becomes much more intricate as the temperature decreases to $T \approx 30$ K (see, for example, the curves for T = 30.9 and 20.1 K in Fig. 5a). In the latter case, the main component of the signal $\rho_H(\varphi) \propto \cos\varphi$ is supplemented with the contribution from the even harmonic $\rho_H(\varphi) \propto \cos 2\varphi$ (see Fig. 5a) in the entire range of magnetic fields used in the present study. The presence of the component from the even harmonic, which considerably complicates the behavior of the angular dependences and the analysis of the Hall effect in FeSi can be traced most clearly in the temperature range 10–20 K (see Fig. 5a), corresponding to the close neighborhood of the sign-inversion temperature $T_{inv}^H \approx 12.5$ K of the main component of the Hall signal [25].

Concluding this subsection, we note that the emergence of even harmonics in the angular dependences of

the Hall resistance $\rho_H(\varphi) \propto \cos 2\varphi$ is connected in some cases with the addition to quantity $\rho_H(T, H, \varphi)$ of the contribution from the even (in magnetic field) magnetoresistance due to the asymmetric arrangement of the Hall contacts on the sample. To estimate the effect of the nonequipotential nature of the Hall contacts, the angular dependences of the magnetoresistance $\rho(\varphi, H, T)$ of FeSi were analyzed simultaneously with the Hall measurements. The results of measurements of $\rho(\varphi, H, T)$ proved that the amplitude of even harmonics in the dependence $\Delta\rho/\rho = f(\varphi)$ does not exceed 0.5% in the entire range of magnetic field under study up to 70 kOe. At temperatures above 7 K, the singularities on the angular dependence of $\Delta\rho/\rho$ are suppressed completely (see also [23]). Thus, the results of our experiments have made it possible to completely eliminate the effect of the conventional resistive component emerging due to the nonequipotentiality in the arrangement of the Hall contacts on the form and the variation of the Hall resistance $\rho_H(T, H, \varphi)$ for all FeSi samples studied by us here.

3.3. Magnetization and Microwave Magnetoabsorption

In studying the Hall effect in a magnetically ordered state of the substance, the spontaneous magnetization of the magnetic subsystem leads to the emergence of an additional (anomalous) component in the Hall resistance, which is proportional to the intrinsic magnetic field [31]. To obtain information required for separating the normal and anomalous contributions to the Hall effect and to refine the peculiarities in the magnetic properties and phase transitions while constructing the low-temperature magnetic *H–T* phase diagram of FeSi, we measured the magnetization and the microwave magnetoabsorption of the samples.

Using the results of precision measurements of magnetization $M(H, T)$ in FeSi made for magnetic fields $H \le 120$ kOe at fixed temperatures 1.6 K $\le T \le$ 50 K, we calculated the dependences of the differential magnetic susceptibility

$$\chi_d(H, T) = \left.\frac{dM}{dH}\right|_T = f(T),$$

shown in Fig. 7 with the help of the numerical differen-

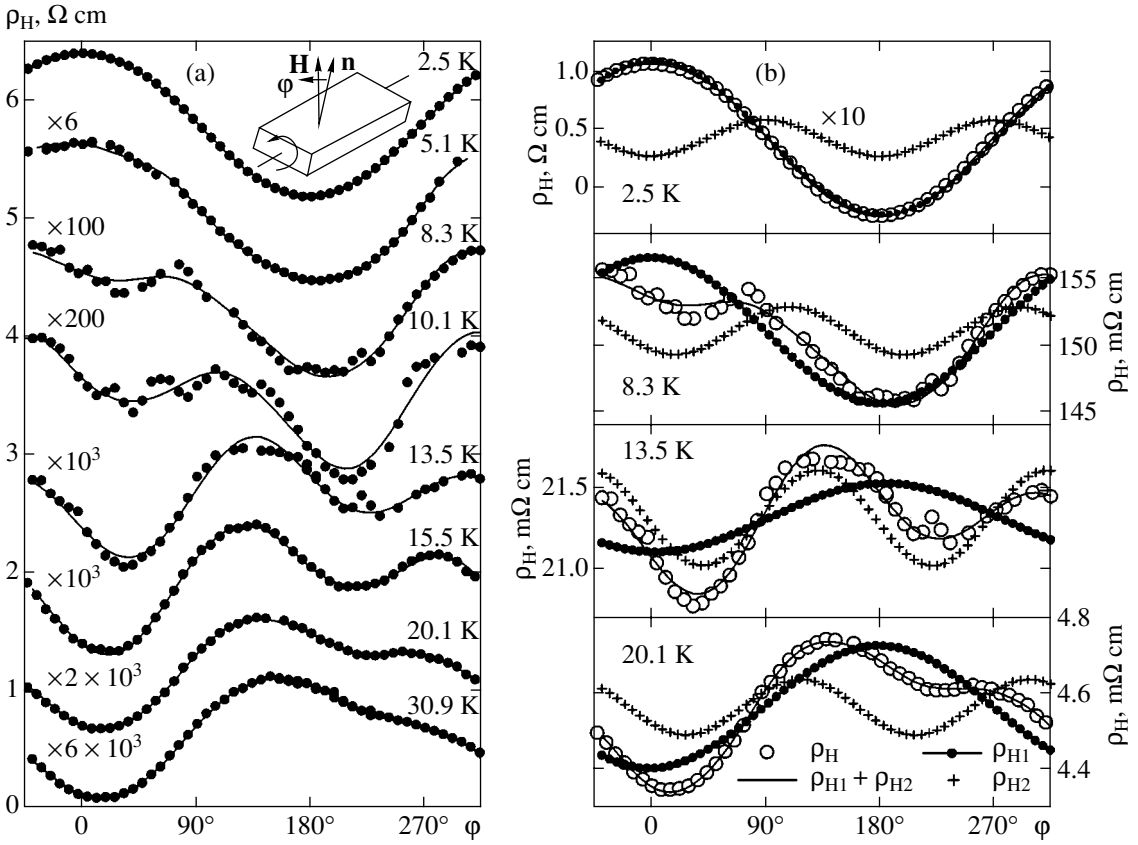


Fig. 5. (a) Angular dependences of the Hall resistance $\rho_H(\varphi)$ for various temperatures in a magnetic field $H = 59.3$ kOe approximated by expression (1) taking into account the contribution from the first ($\rho_{H1} \propto \cos\varphi$) and second ($\rho_{H2} \propto \cos 2\varphi$) harmonics. For convenience of comparing the amplitudes of the Hall effect, the factors for comparing the values of ρ_H on the absolute scale are given on the left of the curves. The inset illustrates schematically the measurement of the Hall resistance $\rho_H(\varphi)$ with the sample rotated relative to the magnetic field **H** (**n** is the normal to the sample surface). (b) Separation of the contributions to $\rho_H(\varphi)$ using relation (1) (see text) for various temperatures at $H = 59.3$ kOe (ρ_H are experimental data, ρ_{H1} is the contribution from the main component, and ρ_{H2} is the contribution from the second harmonic).

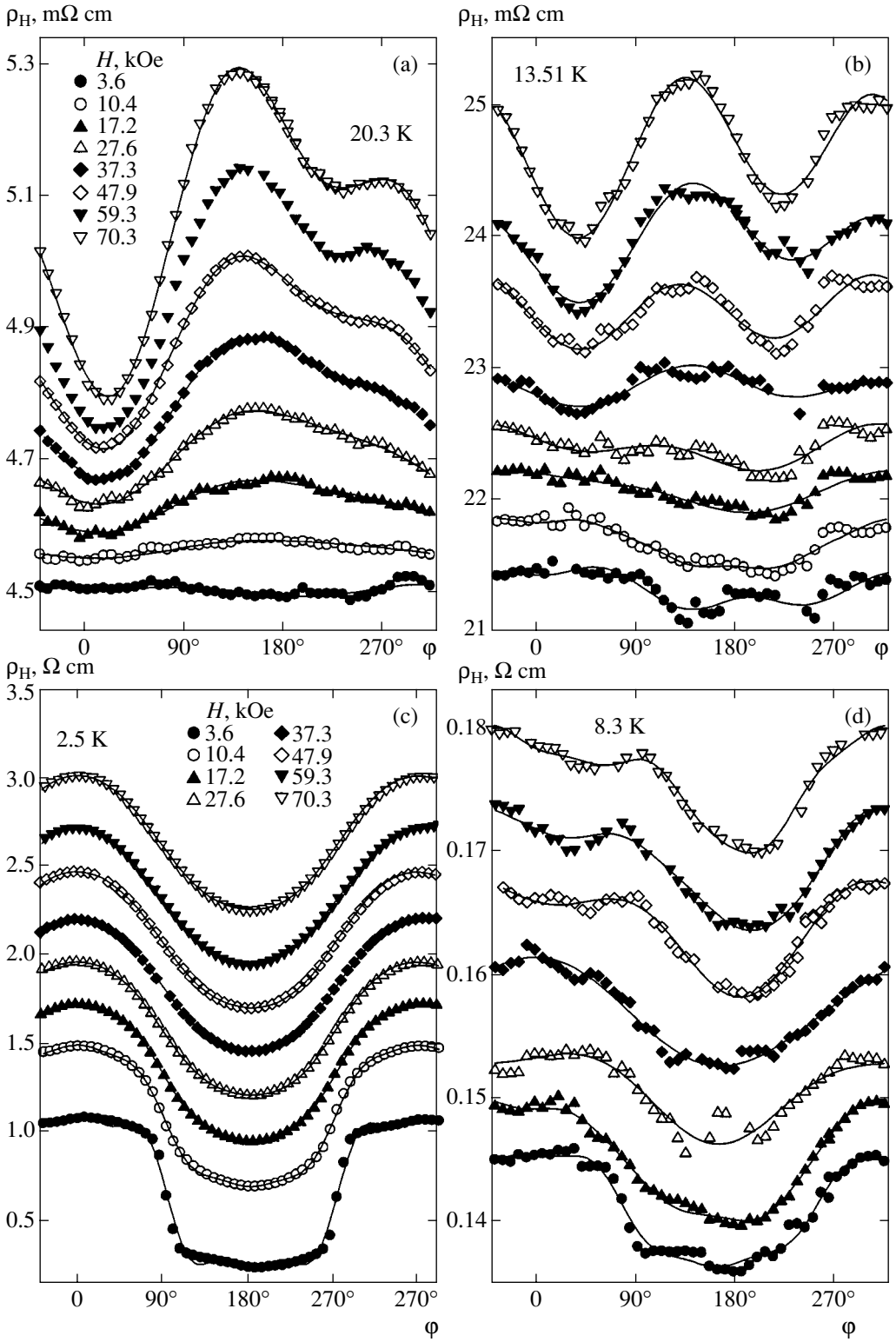


Fig. 6. Angular dependences of the Hall resistance $\rho_H(\varphi)$ of FeSi in magnetic fields up to 75 kOe for temperatures T = 20.3 K (a), 13.5 K (b), 2.5 K (c), and 8.3 K (d).

tiation method. It can be seen from the figure that the magnetic susceptibility considerably decreases in absolute value with increasing external magnetic field: the value of χ_d at T_0 = 1.5 K in field H = 110 kOe decreases approximately by a factor of 2.5 (see also the inset to Fig. 7). In our earlier experiments [23, 32], it was shown that, at low temperatures $T \le 50$ K, the decisive contribution to χ_d comes from the Pauli paramagnetic

susceptibility of the system of heavy fermions (spin polarons). The decrease in the absolute values of χ_d corresponds to partial suppression of the singularity in the electron density of states (many-particle resonance) at the Fermi level in a strong magnetic field $H \le 120$ kOe. Since the exchange field of spin-polaron states $H_{ex} \approx 350 \pm 100$ kOe determined in [23, 32] during measurements of the anomalous Hall effect and magnetization in FeSi is comparable to the limiting value of $H \approx 120$ kOe used in this study, the observed decrease in the value of $\chi_d(T)$ with increasing H (Fig. 7) appears quite substantiated.

Analysis of the family of the $\chi_d(T)$ curves (see Fig. 7) shows that the form of the $\chi_d(T)$ dependence changes in the temperature range $T < T_m \approx 7$ K, corresponding to the mictomagnetic state of FeSi in the field $H \approx 35$ kOe. The decrease in the differential susceptibility observed upon an increase in temperature in the range of weak fields ($H < 35$ kOe) is replaced by the increase in χ_d with temperature for $H > 35$ kOe (see Fig. 7). As a results, the $\chi_d(T)$ curves acquire a peak at a temperature T^{χ}_{max} in fields above 35 kOe, which is displaced upwards along the temperature scale upon an increase in the magnetic field up to values of $T^{\chi}_{max} \approx 12$ K at $H = 120$ kOe (see Fig. 7). The nonmonotonic temperature dependence of the differential susceptibility is obviously responsible for the "intersection" of the field dependences $\chi_d(H, T < T_m \approx 7\ \text{K})$ in the vicinity of $H \approx 35 \pm 5$ kOe (see the inset to Fig. 7). These anomalies in the magnetic properties of FeSi, which are observed near the temperatures T_m and T_C, will be used in Section 4 in analyzing the peculiarities of the low-temperature magnetic H–T phase diagram of FeSi.

The results of investigation of the microwave magnetoabsorption at temperature $T < 6$ K corresponding to the mictomagnetic state of FeSi (SG phase in Fig. 1c) are presented in Fig. 8. It can be seen that the curves describing magnetoabsorption of electromagnetic radiation in the millimeter wavelength range exhibit a hysteresis in the vicinity of $H \approx 15$ kOe (Fig. 8a) and $H \approx 35$ kOe (Fig. 8b). It is interesting to note that the values of the fields corresponding to anomalies in the microwave magnetoabsorption in FeSi correlate well with the anisotropy field $H_{an} \approx 12 \pm 2$ kOe of ferromagnetic nanoclusters [23] and with the field $H \approx 35 \pm 5$ kOe corresponding to singularities on the differential susceptibility curves $\chi_d(H, T < T_m \approx 7\ \text{K})$ (see the inset to Fig. 7). In the latter case, the coincidence of the characteristic magnetic fields and the presence of hysteretic anomalies apparently indicate the field-induced magnetic transition in the low-temperature phase of FeSi.

Concluding the section, we must emphasize that the high precision of the microwave magnetoabsorption method makes it possible to detect weak hysteretic anomalies at a level of $\Delta P/P \approx 0.2\%$ (Fig. 8). The small magnitude of the effect apparently indicates a relatively

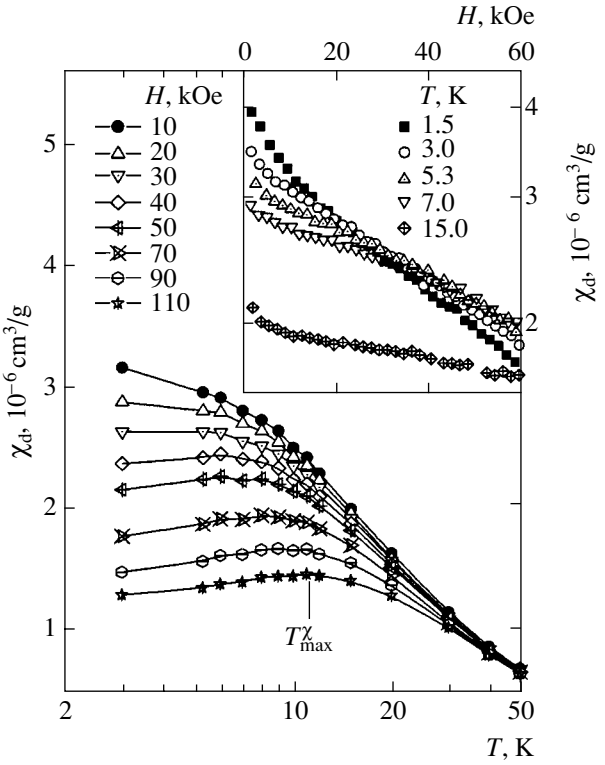


Fig. 7. Temperature dependences of the differential susceptibility $\chi_d(T)$ of FeSi in a magnetic field up to 120 kOe. The inset shows the field dependences $\chi_d(H)$ in the temperature range 1.5 K $\le T \le$ 15 K corresponding to the formation of ferromagnetic microregions in the FeSi matrix.

low concentration of magnetic microregions determining the hysteresis of microwave magnetoabsorption in FeSi. In such a situation, the absence of hysteretic anomalies in the field dependences of differential susceptibility in the field interval 15–35 kOe (see Fig. 7) can be attributed to the insufficient accuracy of the numerical differentiation method (1–5%), which practically rules out the possibility of quantitative analysis of the magnetization reversal effects in a system of magnetic microscopic regions in the FeSi matrix on the basis of the results of magnetic measurements. At the same time, simultaneous analysis of magnetic and magnetooptical properties makes it possible to establish quite reliably the position of singularities associated with low-temperature magnetic transitions in the FeSi matrix in magnetic fields up to 70 kOe.

4. DISCUSSION

4.1. Separation of Contributions to the Hall Effect

In analyzing the results of measurements of the angular dependences of the Hall resistance of FeSi samples in the temperature range 4.2–30 K (Figs. 5 and 6), we used the representation

$$\rho_H(\varphi, T, H) = \rho_{H0} + \rho_{H1}\cos\varphi + \rho_{H2}\cos(2\varphi - \Delta\varphi), \quad (1)$$

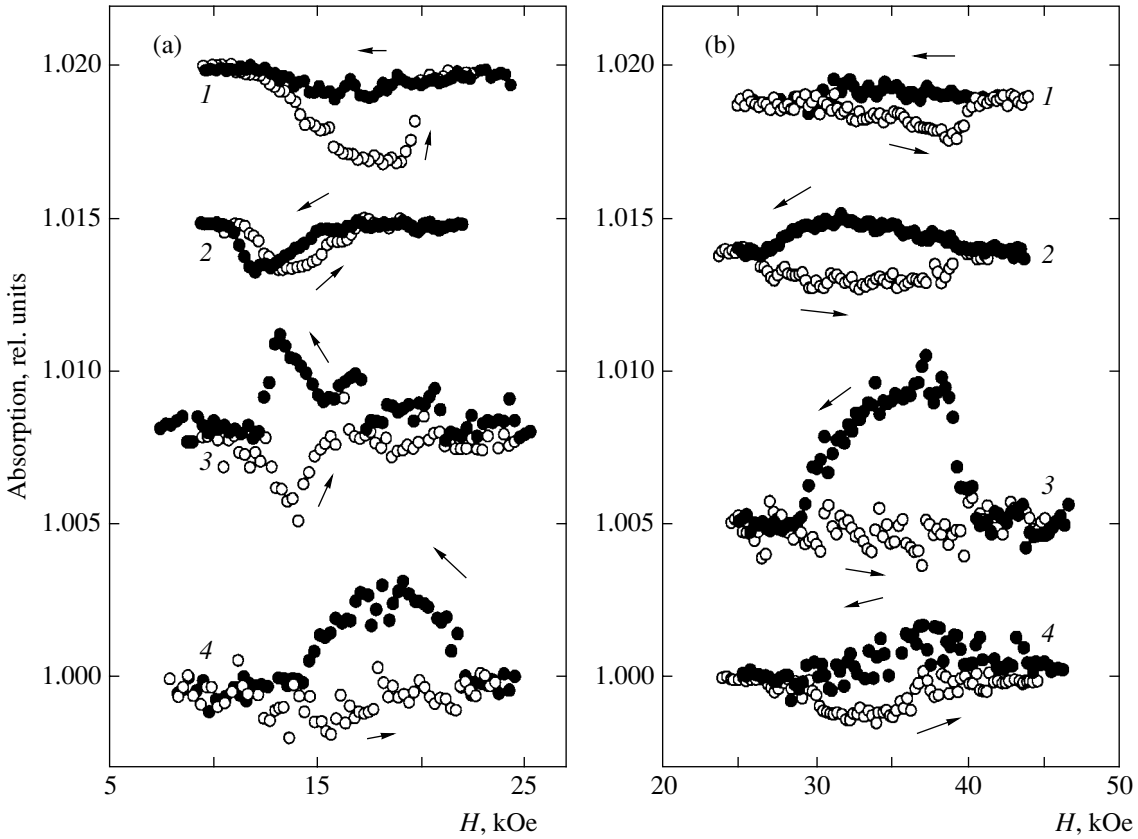


Fig. 8. Anomalies in the microwave magnetoabsorption of FeSi in the vicinity of the anisotropy field of ferromagnetic microregions, $H_{an} \approx 15$ kOe (a) and the boundary of the mictomagnetic ground state $H_m \approx 35$ kOe (see Section 4.3) (b) for various values of the electromagnetic radiation frequency and temperature: (*1*) $\nu = 37.55$ GHz, $T = 4.16$ K; (2) $\nu = 37.55$ GHz, $T = 4.42$ K; (3) $\nu = 37.55$ GHz, $T = 5.16$ K; and (4) $\nu = 46.33$ GHz, $T = 5.69$ K. The arrows indicate the direction of magnetic field variation.

which takes into account, in addition to the main component ρ_{H1} (odd in the magnetic field) and the constant shift ρ_{H0} emerging due to asymmetry of the Hall contacts, the second harmonic contribution ρ_{H2} with a shift $\Delta\varphi$ relative to the first harmonic. Figure 5b shows for visualization several examples of separation of the contributions to $\rho_H(\varphi, T, H)$ for the experimental results obtained from measurements in a magnetic field $H =$ 59.3 kOe at different temperatures. The results shown in Fig. 5b lead to the conclusion about the sign inversion of the main component of the Hall resistance ρ_{H1}; with decreasing ρ_{H1}, contributions ρ_{H1} and ρ_{H2} become comparable in the order of magnitude in a large neighborhood of the inversion temperature T_{inv}^{H} and their competition completely determines the behavior of the Hall signal. In a small neighborhood of the sign-inversion temperature for the main contribution ρ_{H1}, a considerable dependence of T_{inv}^{H} on the external magnetic field is observed in the conditions when the component from the even harmonic is predominant in ρ_H. It is this circumstance that is apparently responsible for the difficulties in the measurements and interpretation of the Hall effect in the temperature range 4.2–30 K, which were noted in [15]. Indeed, when the Hall measurements are performed in FeSi using the conventional scheme with recording of the Hall voltage for two opposite (perpendicular to the plane of the sample) directions of magnetic field H corresponding to $\varphi = 0$ and $\varphi = 180°$ in Figs. 5 and 6, a complex nonmonotonic (in magnetic field and temperature) behavior of the Hall resistance in the temperature range 4.2–30 K can be expected.

In such a situation, the most visual pattern of variation of the components in the Hall effect in FeSi can be obtained from an analysis of the amplitudes of contributions ρ_{H1} and ρ_{H2} as well as the phase shift $\Delta\varphi$ determined from the experimental data (see Figs. 5 and 6) using relation (1). Figures 9 and 10 show the field dependences of the main (ρ_{H1}) and field-even (ρ_{H2}) contributions to the Hall resistance; the inset to Fig. 10 shows the variation of parameter $\Delta\varphi$ with temperature. It can be seen from Fig. 9 that the main (positive in sign and linear in the magnetic field) component ρ_{H1} of the Hall resistance increases in amplitude upon a decrease in temperature at $T < 40$ K. In the vicinity of the Curie temperature $T_C \approx 15$ K, in addition to the linear positive component $\rho_{H1}^{SP}(H)$ corresponding to the transport in

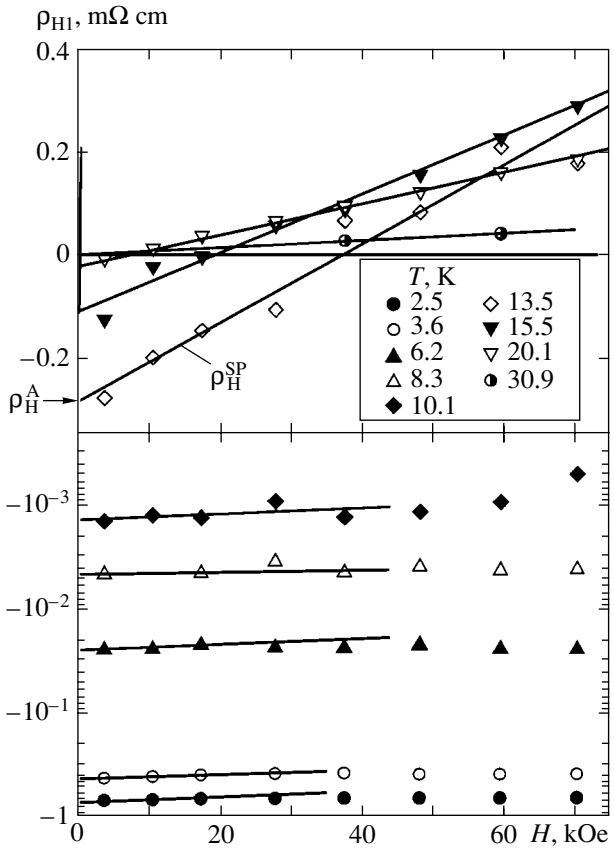


Fig. 9. Field dependences of the main component of the Hall resistance ρ_{H1} in a magnetic field up to 75 kOe. For the dependence $\rho_{H1}(H)$ at $T = 13.5$ K, the separation of the normal $\rho_H^{SP} = R_H^{SP} H$ and the anomalous $\rho_H^A = R_H^A H_{ex}$ contributions to the main component of the Hall resistance is illustrated.

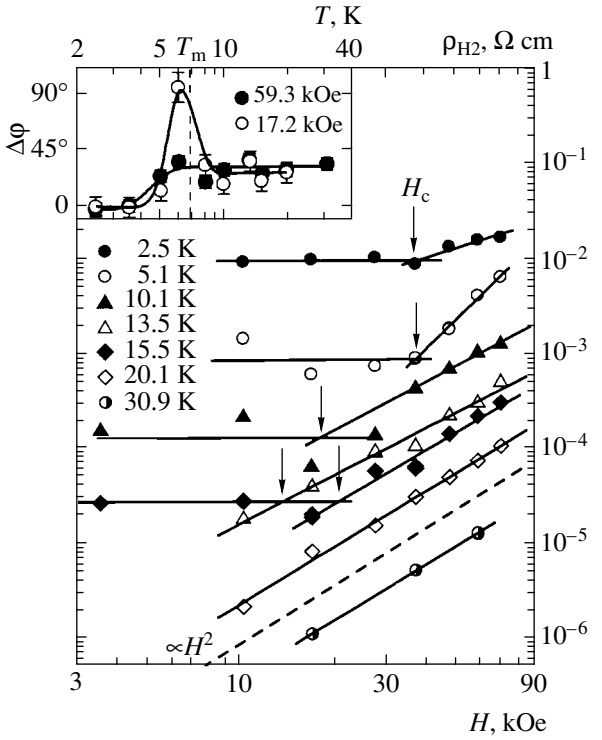


Fig. 10. Field dependences of the second harmonic of the Hall resistance ρ_{H2} represented in the log–log coordinates from the results of measurements in a magnetic field up to 75 kOe in the temperature range 2–32 K. The dashed line shows the quadratic dependence of the form $\rho_{H2} \propto H^2$. The inset shows the temperature dependences of the phase shift $\Delta\varphi$ between the first and the second harmonics (see text) in magnetic fields of 17.2 and 59.3 kOe.

spin-polaron states [23], the quantity ρ_{H1} acquires an anomalous negative contribution ρ_{H1}^A independent of H (see Fig. 9). A further decrease in temperature leads to a substantial increase in the absolute value of the anomalous component ρ_{H1}^A of the Hall resistance (by more than five orders of magnitude upon a decrease in temperature from 20 to 1.8 K), this contribution becoming predominant at temperatures $T \leq 11$ K (see Fig. 9). Thus, analysis of the available experimental data leads to an unambiguous conclusion that the low-temperature sign inversion of the main component in the Hall signal in FeSi is determined by the condition of the equality of the amplitudes of the anomalous negative (ρ_{H1}^A) and the normal positive (ρ_{H1}^{SP}) contributions to $\rho_{H1}(H, T)$.

Under the conditions of the substantial increase in the anomalous contribution to the Hall effect, we can naturally expect a displacement of the sign-inversion temperature T_{inv}^H of the first harmonic ρ_{H1} of the Hall resistance in a magnetic field. Figure 11 shows the temperature dependences of quantities ρ_{H1} and ρ_{H2} for values of the external magnetic field at the beginning (H = 17.2 kOe) and at the end (H = 70.3 kOe) of the range of variation of H used in this research. For the results presented in Fig. 11, the shift of the values of T_{inv}^H with H increasing from 17.2 to 70.3 kOe is approximately 3 K. In turn, the absence of a noticeable dependence of the main component ρ_{H1} of the Hall resistance on the external field $H \leq 70$ kOe for temperatures below 10 K (see Fig. 11) served, in our opinion, as an additional argument in favor of the decisive role of the anomalous contribution to the Hall effect in this temperature range.

The field dependences of the second harmonic $\rho_{H2}(H)$ of the Hall resistance shown in Fig. 10 for the paramagnetic phase at $T > T_C \approx 15$ K are characterized by a nearly quadratic dependence $\rho_{H2}(H) \propto H^2$. In the vicinity of the Curie temperature $T \approx T_C$, the $\rho_{H2}(H)$ curve acquires a segment with $\rho_{H2}(H) = \text{const}$ in the region of weak fields; above this segment, for a certain critical value of the magnetic field H_c, this component of the Hall resistance again exhibits a nearly quadratic asymptotic behavior. With decreasing temperature, the

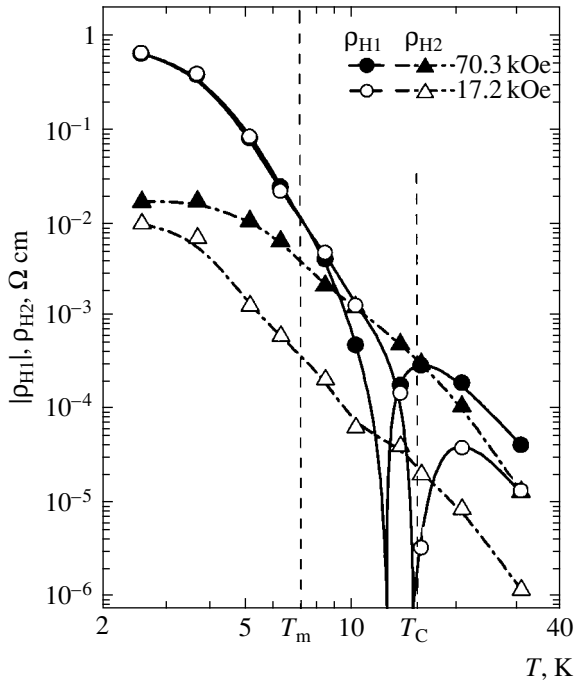


Fig. 11. Temperature dependences of the first $\rho_{H1}(T)$ and second $\rho_{H2}(T)$ harmonics of the Hall resistance in magnetic fields of 17.2 and 70.3 kOe.

point on the $\rho_{H2}(H)$ curves at which the regime changes at H_c is displaced upwards on the H scale; at temperatures $T < T_m \approx 7$ K corresponding to the mictomagnetic state in the FeSi matrix, field H_c attains a constant value of $H_c \approx 35$ kOe.

Another parameter that also experiences noticeable changes upon a transition at $T < T_m \approx 7$ K to the state with magnetic moments of ferromagnetic nanoclusters frozen in the FeSi matrix is the phase shift $\Delta\varphi$ between the first and second harmonics; the temperature dependence of this parameter is shown in the inset to Fig. 10. The data in the inset show that the value of $\Delta\varphi$ at $T > T_m \approx 7$ K amounts approximately to 30°. As the temperature decreases in the immediate vicinity of the transition to the mictomagnetic state at $T_m \approx 7$ K, the phase shift abruptly drops to zero. The temperature dependence of the second harmonic of the Hall resistance (see Fig. 11) indicates that the $\rho_{H2}(H)$ curves repeat on the whole the behavior of the main component of the Hall signal. However, since $\rho_{H2}(H)$ is an even component of the Hall resistance $R_H(T)$ in magnetic field, instead of the sign inversion in the vicinity of T_C, the $\rho_{H2}(H, T)$ curves exhibit only a weak singularity in the form of a point of inflection in the temperature range $T_m < T < T_C$ (see Fig. 11).

It is interesting to note that a similar magnetic-field-even contribution to the Hall effect (second harmonic in the angular dependences) was observed while studying the galvanomagnetic characteristics of the compound $CeAl_2$ with heavy fermions [33]. In this compound, a decrease in the temperature in the vicinity of the transition to the antiferromagnetic state ($T < T_N \approx 3.85$ K) was accompanied by a substantial increase in the amplitude of the anomalous component of $\rho_{H2}(H, T)$; the emergence of the corresponding magnetic component of the Hall coefficient $R_H^{am} = \rho_{H2}(H, T)/H$ was attributed by the authors [33] to magnetization reversal processes in magnetic domains and to the existence of short-range ferromagnetic correlations in the $CeAl_2$ matrix in the close neighborhood of the antiferromagnetic phase. At the same time, a monotonic variation of $\rho_{H2}(H, T)$ in the immediate vicinity of the temperature $T_C \approx 15$ K corresponding to the formation of ferromagnetic microregions along with the existence of magnetic-field-even component in the Hall effect in FeSi at temperatures substantially higher than T_C (see Fig. 11) apparently rules out the direct correlation of this component of the Hall signal with the formation of ferrons in FeSi. Thus, the origin of this even anomalous component in the Hall signal remains unclear and further investigations are required to clarify the origin of the second harmonic $\rho_{H2}(H, T)$ of the Hall resistance in FeSi and other compounds with strong electron correlations.

Returning to the discussion of the main magnetic-field-odd component ρ_{H1} of the Hall resistance in FeSi, it should be emphasized that the anomalous contribution ρ_H to the Hall effect is observed at temperatures below $T_C \approx 15$ K, at which ferromagnetic nanoclusters are formed [23]. In contrast to the behavior predicted for traditional magnetic metals and semiconductors ($R_H^A \longrightarrow 0$ as $T \longrightarrow 0$ [34]), the absolute value of the anomalous component of the Hall resistance in FeSi increases with decreasing temperature by five orders of magnitude, attaining values of $\rho_H^A \approx 1$ Ω cm (see Fig. 9) and does not exhibit a tendency to decrease even at ultralow temperatures $T \approx 0.05$ K [15]. In addition, the difference in the signs of the spin-polaron $\rho_H^{SP}(H, T)$ and anomalous ($\rho_H^A(T)$) contributions to the Hall resistance of FeSi, which correspond to the hole ($\rho_H^{SP}(H, T) > 0$) and the electron ($\rho_H^A(T) < 0$) types of conduction, is not typical of traditional magnetic materials either. An analogous difference in the signs of the anomalous and normal Hall coefficients was observed in the study of galvanomagnetic characteristics of manganites of the family La(Ca, Sr)MnO_3 in a broad neighborhood of the Curie temperature corresponding to the regime of transport over spin-polaron states (see, for example, [35, 36]). To interpret the nontrivial temperature dependence of the Hall coefficients in manganites, a model was proposed on the basis of the assumption concerning the phase change in the electron wavefunction under the effect of the local magnetic surroundings [37]. In such a situation, the motion of a charge carrier in a

system of noncomplanar spins is accompanied by the emergence of an additional phase shift (Berry's phase), which is similar to the emergence of the effective intrinsic magnetic field for the charge carrier [38]. The allowance for the additional geometrical phase under the conditions of a strong spin–orbit interaction makes it possible to qualitatively describe the difference in sign and the temperature dependence of the anomalous Hall coefficient in manganites [37]. At the same time, direct application of the model [37] taking into account the effect of the region polarized by the charge carrier on the charge transfer processes in a magnetic field seems hardly possible when the Hall effect is analyzed in the regime of spin-polaron transport in FeSi.

Taking into account the above arguments, we obtained a quantitative description of the Hall effect in FeSi using the common procedure of separation of the Hall resistance components in ferromagnets (see, for example, [31]). We represented the set of experimental data in terms of the "normal" (spin-polaron) R_H^{SP} and anomalous R_H^A Hall coefficients in the form

$$\begin{aligned}\rho_{H1}(H, T) &= \rho_H^{SP}(H, T) + \rho_H^A(T) \\ &= R_H^{SP}(H, T)H + R_H^A(T)H_{ex},\end{aligned} \quad (2)$$

where

$$\rho_H^{SP}(H, T) = R_H^{SP}(H, T)H, \quad \rho_H^A(T) = R_H^A(T)H_{ex},$$

and $H_{ex} = 350 \pm 100$ kOe is the exchange field at the charge carrier [23, 27]. It should be emphasized that, in contrast to our earlier publication [23], instead of the term $R_H^A(T)M(T)$ traditionally used for describing the anomalous Hall effect [31], the right-hand side of formula (2) contains the representation $\rho_H^A(T) = R_H^A(T)H_{ex}$, in which the temperature dependence of magnetization, $M(T)$, is directly contained in $R_H^A(T)$. In our opinion, such an approach is justified since the structure of the magnetically ordered phase (ferromagnetic nanoclusters in a small concentration in the weakly magnetic FeSi matrix) requires the replacement of the volume magnetization M of the material in the expression for ρ_H^A by its local value $M_{loc}(T)$ determined by the effective magnetic field at the charge carrier in the upper (t_{2g}) Hubbard band. An example of representation of the experimental data on $\rho_{H1}(H, T)$ in terms of model (2) is given in Fig. 9.

The approach described above was applied here to calculate the Hall coefficients R_H^A and R_H^{SP} for magnetic fields of 17.2 and 70.3 kOe (Fig. 12a). Using the values of R_H^A and R_H^{SP} (Fig. 12a), we can estimate the

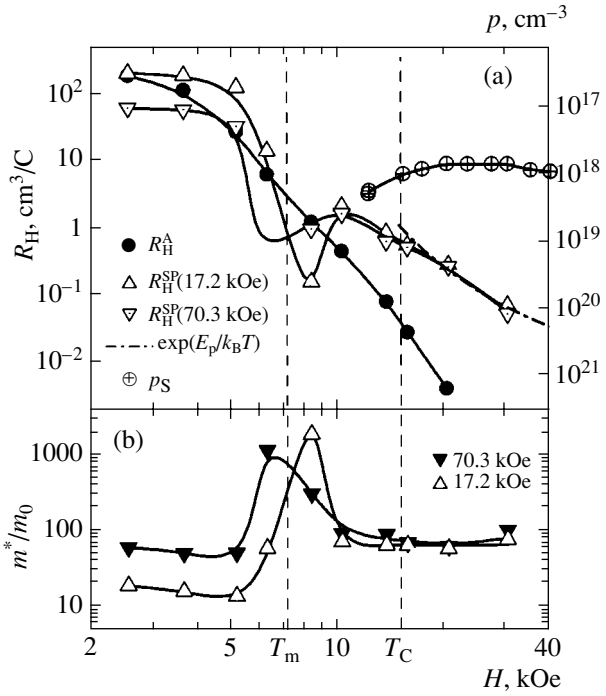


Fig. 12. (a) Temperature dependences of the effective concentration $p = (R_H e)^{-1}$ obtained for the normal $R_H^{SP}(H, T)$ and anomalous $R_H^{SP}(T)$ Hall coefficients (see formula (2)) as well as the concentration of charge carriers $p_S = N_{Fe}\exp(|e|S/k_B)$ calculated from the thermo-emf coefficient S in the regime of the spin-polaron transport [23]. The dot-and-dash curve shows the activation asymptotic behavior of the Hall coefficient $R_H^{SP}(T) \propto \exp(E_p/k_BT)$ with activation energy $E_p \approx 6$ meV at $T > 15$ K [23]. (b) Variation of the effective mass of charge carriers in FeSi in magnetic fields of 17.2 and 70.3 kOe.

local magnetization $M_{loc}(T)$ of ferromagnetic microregions in FeSi at temperatures $T \le 20$ K using the relation $M_{loc} = \rho_H^A/4\pi R_H^{SP}$ proposed in [23] and based on application of the model of two groups of charge carriers for describing the anomalous Hall effect. It should be emphasized that an analogous approach is used in the literature for determining the equivalent field $H_A = \rho_H^A/R_H$, for which the magnitude of the normal Hall effect coincides in the linear approximation with the anomalous Hall effect [34]. The temperature dependence of the effective field $4\pi M_{loc}$ (Fig. 13) shows that a decrease in the temperature leads to a considerable increase in the value of local magnetization. In this case, the increase in the value of M_{loc} in the immediate vicinity of $T_C \approx 15$ K turns out to be a much slower process as compared to the case of "classical" band ferromagnets (in Fig. 13, the magnetization curves for ferromagnetic metals Fe and Ni, as well as the theoretical

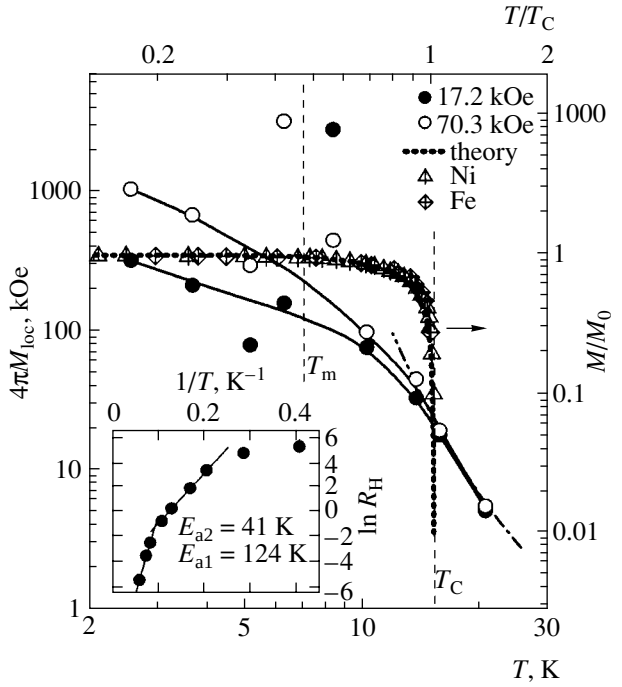


Fig. 13. Temperature dependences of the effective field $4\pi M_{loc} = \rho_H^A / R_H^{SP}$ represented in the log–log coordinates from the results of measurements in magnetic fields of 17.2 and 70.3 kOe. The dot-and-dash curve shows the activation asymptotic behavior of parameter $M_{loc}(T)$ (see text). For comparison, the experimental dependences of spontaneous magnetization $M/M_0 = f(T/T_C)$ for iron and nickel in normalized coordinates are given together with the theoretical dependence $M/M_0 = f(T/T_C)$ calculated in the mean field approximation, where $M_0 = M$ (T = 0 K) [34]. The inset shows the activation asymptotic forms of the anomalous Hall coefficient R_H^A.

dependence of spontaneous magnetization of a ferromagnet at $T \le T_C$ [39], are also shown for comparison).

Analysis of the results presented in Fig. 13 shows that, instead of the critical behavior in the vicinity of the Curie temperature, nearly activation-type dependences $M_{loc}(T)$ (Fig. 13) and $R_H^A(T)$ (inset to Fig. 13) are observed for the system of ferromagnetic clusters in the FeSi matrix. In spite of the approximate nature of numerical estimates, the values of the activation energy characterizing the temperature dependences of M_{loc} ($E_a \approx 6.5$ meV) and R_H^A ($E_{a1} \approx 12$ meV and $E_{a2} \approx 4$ meV) are in fairly good agreement with the binding energy of spin-polaron states in the FeSi matrix, $E_p \approx 6$ meV [23]. It should also be noted that the essentially nonmonotonic behavior of the dependences $R_H^{SP}(T)$ (see Fig. 12a) and $M_{loc}(T) = \rho_H^A / 4\pi R_H^{SP}$ (see Fig. 13) in the vicinity of $T_m \approx 7$ K should be associated, in our opinion, with the noticeable enhancement of carrier scattering by the inhomogeneities appearing in the FeSi matrix in the vicinity of the transition to the mictomagnetic state. In addition, the value of the magnetization $4\pi M_{loc} \approx 500 \pm 200$ kOe at T = 2.5 K coincided to within the experimental error with the value of the exchange field at a carrier in the upper Hubbard band, $H_{ex} = 350 \pm 100$ kOe, determined in [23] for FeSi.

The temperature dependences of the normal (R_H^{SP}) and anomalous (R_H^A) Hall coefficients obtained in this study (see Fig. 12a) can also be analyzed in terms of the effective concentration of charge carriers defined as $p = (R_H e)^{-1}$ (see the ordinate axis on the right in Fig. 12a). In our opinion, it would be most interesting to compare the behavior of the effective Hall concentration $p_H(T) = (R_H^{SP} e)^{-1}$ with the concentration of charge carriers estimated from the measurements of the temperature dependence of the thermo-emf coefficient [23]. It should be recalled that the thermo-emf coefficient in the regime of strong Coulomb correlations is determined by the reduced concentration of charge carriers in the Hubbard band [40]. Thus, in the temperature range $T_C < T < 100$ K corresponding to the transport over spin-polaron states at the Fermi level, the data on thermo-emf in FeSi [23] enable us to directly estimate the charge carrier concentration by the Hickes formula $p_S = N_{Fe}\exp(|e|S/k_B)$ [40]. The obtained values of p_S are practically independent of temperature in the interval 15–50 K: $p_S \sim 10^{17}$–10^{18} cm^{-3} (see Fig. 12a). It should also be noted that a similar estimate of the charge carrier concentration ($p \sim 3 \times 10^{18}$ cm^{-3}) was obtained while studying the low-temperature specific heat of FeSi [41]. A comparison of the data presented in Fig. 12a clearly demonstrates that the absolute value and the activation nature of the dependence of the Hall parameter $p_H(T) = (R_H^{SP} e)^{-1}$ substantially differ from the behavior of concentration p_S estimated from the results of thermo-emf measurements [23]. A natural explanation of the noted difference in the parameters $p_H(T)$ and $p_S(T)$ can be obtained in the model of carrier transport over small-radius (~10 Å) spin-polaron states, which is characterized by the activation dependence of the Hall coefficient R_H^{SP}. The activation energy for the dependence $R_H^{SP}(T)$ is determined by the spin-polaron potential $E_p \approx 6$ meV [23] in contrast to the thermo-emf coefficient that varies with temperature only slightly and can be recalculated by the Hickes formulas to the actual concentration of charge carriers in the Hubbard bands ($p_S(T) \sim 10^{17}$–10^{18} cm^{-3}; see Fig. 12a).

Another interesting (in our opinion) comparison of the effective concentration parameters can be carried out for temperatures $T < T_m \approx 7$ K, at which the values of $p_A = (R_H^{SP} e)^{-1}$ and $p_H(T) = (R_H^{SP} e)^{-1}$ are found to be close. We believe that, in this temperature range corre-

sponding to the formation of the mictomagnetic state in FeSi, the data presented in Fig. 12a indicate that the concentration of heavy fermions and nanosize ferromagnetic regions in the FeSi matrix are comparable. The decrease in the values of effective concentration of charge carriers as compared to the concentration $p_S = 5.8 \times 10^{17}$ cm^{-3} of spin-polaron states in FeSi in the paramagnetic state ($T < T_C$) should apparently be attributed to the "actuation" of the interaction between ferromagnetic nanosize regions, which leads to their amalgamation into clusters accompanied by the transition to a state with frozen spins (SG state in Fig. 1c).

Thus, the separation of the channels forming the main component of the Hall signal ($R_H^{SP}(H, T)$ and $R_H^A(T)$ in Figs. 9 and 12) enables us to draw the following conclusions. The positive Hall coefficient $R_H^{SP}(H, T)$ in FeSi is apparently associated with the contribution from spin-polaron states forming a many-particle resonance at the Fermi level (structure I in Fig. 1). This "resonant component" is analogous to the well-known anomalous positive Hall effect observed in heavy-fermion systems based on rare-earth elements (see, for example, 42, 43]. In our opinion. it is the formation of a many-particle resonance in the density of states in the vicinity of the Fermi energy E_F in FeSi that leads to the formation of a positive spin-polaron component of the Hall coefficient at temperatures $T <$ 100 K. The activation-induced increase in the value of $R_H^{SP}(T) \propto \exp(E_p/k_BT)$ with decreasing temperature (see Fig. 12a and [23, 25]) makes it possible to estimate the binding energy $E_p \approx 6$ meV of spin-polaron states in the FeSi matrix (structure I in Fig. 1). As the density of states for the many-particle resonance increases at $T <$ 20 K, a transition to the coherent mode of Hubbard fluctuations occurs between the e_g and t_{2g} states of the 3*d* band of iron and to the formation (at $T = T_C \approx 15$ K) of ferromagnetic nanosize regions (ferrons) from spin polarons. Taking into account the estimates of the energy $U \approx 270$ meV of Hubbard correlations and the electron density of states $N \geq 20$ states/(eV unit cell) obtained in our previous publication [23], we have $UN > 1$ for a ferromagnetism criterion similar to the Stoner criterion; for FeSi, we obtain $UN \geq 5$. Apparently, the fulfillment of the condition $UN \geq 5$ enables us to explain the emergence of spontaneous magnetization and, as a result, the emergence of the anomalous ferromagnetic component $R_H^A(T)$ of the Hall effect in FeSi at low temperatures. The observation of the local magnetization and, accordingly, the anomalous component of the Hall coefficient, $R_H^A(T)$, in a temperature range slightly higher than the Curie temperature $T_C \approx 15$ K (see Fig. 12a; $R_H^A(20\text{ K}) \neq 0$) should probably be attributed to a random distribution of many-particle complexes with a low concentration (of approximately 10^{18} cm^{-3}) in the bulk of a FeSi sample.

4.2. Separation of Contributions to Magnetoresistance

The behavior of the magnetoresistance of FeSi in the temperature range 1.6–30 K turns out to be quite complex and sign-alternating (see Figs. 2–4). In contrast to the main component ρ_{H1} of the Hall resistance, whose sign inversion occurs at temperatures near $T_C =$ 15 K, the sign inversion of the ratio $\Delta\rho/\rho$ (transition from a positive magnetoresistance $\Delta\rho/\rho > 0$ to a negative magnetoresistance $\Delta\rho/\rho < 0$) is observed upon a decrease in temperature in the immediate vicinity of $T_m = 7$ K (Fig. 4). It was noted above that the field dependences of magnetoresistance $\Delta\rho/\rho = f(H)$ at $T > T_m$ exhibit a noticeable deviation from the quadratic dependence $\Delta\rho/\rho \propto H^2$ (see the inset to Fig. 2). It should also be noted that the peak of the positive contribution to the magnetoresistance of FeSi is observed at $T \approx 11$ K, which is substantially lower than the Curie temperature $T_C \approx 15$ K (see Fig. 4).

In our opinion, the most plausible explanation of the effect of positive magnetoresistance at $T > T_m \approx 7$ K can be formulated as follows. Under the conditions of rapid electron density fluctuations (Hubbard correlations at Fe centers), the formation of spin-polaron states at charge carriers occurs due to polarization of the magnetic moments from the nearest neighborhood, which determines the amplitude of the corresponding polaron potential (the depth of the spin-polaron well). In such a situation, additional polarization of nearest neighbors of a charge carrier under the effect of an external magnetic field leads to an increase the depth of the polaron potential and, as a consequence, to a positive contribution to the magnetoresistance of FeSi. In the simplest approximation, an increase in the binding energy of the spin-polaron complex in the paramagnetic FeSi matrix can be taken into account in terms of the change in the magnetostatic energy of the polarized region in a magnetic field [44]:

$$\Delta E = -MH/2, \tag{3}$$

where $M = \chi_{loc} V_{SP} \rho_d H$ is the magnetization of the surroundings polarized by the charge carrier, χ_{loc} and V_{SP} are the local susceptibility and the volume of the polarized region, and $\rho_d = 6.1$ g/cm^3 is the density of FeSi [15]. Using the value of localization radius $r_{SP} \approx 5$ Å of the spin-polaron state [23], we obtain the following expression for the volume of a spin polaron:

$$V_{SP} = 4\pi r_{SP}^3/3 \approx 5.2 \times 10^{-22}\text{ cm}^3.$$

The change in the binding energy of many-particle states in the activation dependence of resistivity

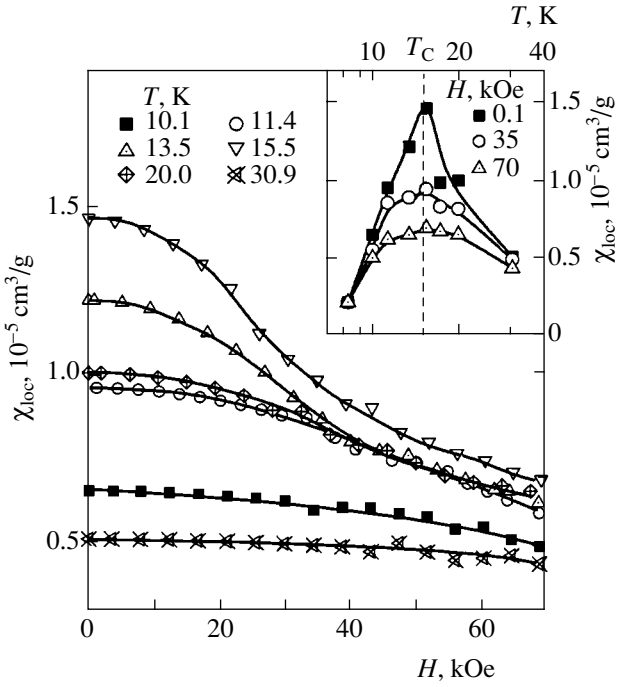


Fig. 14. Field dependences of the local susceptibility χ_{loc} estimated in the framework of model (3)–(5) for various values of temperature in the interval 10–40 K. The inset shows the variation of χ_{loc} with temperature in magnetic fields of 0.1, 35, and 70 kOe.

$\rho(H, T) = \rho_0(T)\exp[E_p(H)/k_BT]$ [23, 25] in a magnetic field can be written in the form

$$E_p(H) = E_p(0) + \Delta E = E_p(0) - \frac{1}{2}\chi_{loc}V_{SP}\rho_d H^2. \quad (4)$$

Under the condition of smallness of parameter $\Delta E/k_BT < 0.03 \ll 1$ (see Figs. 2 and 7), we obtain the following expression for the field dependence of positive magnetoresistance at a fixed temperature:

$$\frac{\Delta\rho}{\rho} = 1 - \exp\left(\frac{\Delta E}{k_BT}\right) \approx \frac{\chi_{loc}V_{SP}\rho_d H^2}{2k_BT}. \quad (5)$$

Thus, in the framework of model (3)–(5), the quadratic term in expression (5) can be used to estimate the local magnetic susceptibility

$$\chi_{loc} = \frac{2k_BT}{H^2V_{SP}\rho_d}\frac{\Delta\rho}{\rho}$$

in the volume V_{SP} occupied by many-particle states of the spin-polaron type in the FeSi matrix.

The results of calculation of the local susceptibility $\chi_{loc}(H, T)$ for the range of temperatures and magnetic fields studied here are shown in Fig. 14. It can be seen that the behavior of field dependences of parameter χ_{loc} correlate well with the variation of the differential magnetic susceptibility $\chi_d(H, T)$ in a magnetic field (see the inset to Fig. 7). For example, as the external magnetic field strength increases, the absolute value of χ_{loc} substantially decreases, the strongest variation of the local magnetic susceptibility in a magnetic field corresponding to temperatures in the immediate vicinity of T_C = 15 K (Fig. 14). At the same time, we must emphasize the difference in the behavior of the differential magnetic susceptibility $\chi_d(H, T)$ averaged over the sample volume and the local susceptibility $\chi_{loc}(H, T)$, which is seen most clearly from a comparison of the temperature dependences of these parameters at fixed values of the magnetic field. The dependences $\chi_{loc}(H, T)$ shown in the inset to Fig. 14 exhibit a peak at $T \sim T_C$ = 15 K, which corresponds to the formation of nanosize ferromagnetic regions in the FeSi matrix. An increase in the magnetic field leads to a considerable suppression of the peak amplitude of χ_{loc}, which is accompanied by blurring of the singularity near T_C = 15 K (see the inset to Fig. 14). It should also be noted that the field dependence of χ_{loc} calculated from the results of measurements of positive magnetoresistance in the temperature range 10–20 K coincide to within the experimental error with those obtained in a magnetic field above 35 kOe (see Fig. 14).

The observed difference in the behavior of $\chi_d(H, T)$ and $\chi_{loc}(H, T)$ can be naturally associated with peculiarities in the formation of the ground magnetic state in FeSi. For example, the bulk magnetic properties of FeSi at $T < 100$ K are determined by the Pauli paramagnetic contribution from many-particle (spin-polaron) resonance emerging in the regime of rapid spin fluctuations at Fe centers, which are accompanied by transitions between Hubbard bands [32]. In turn, parameter $\chi_{loc}(H, T)$ turns out to be connected with the exchange enhancement of magnetization due to polarization of Fe centers in the immediate vicinity of a change carrier and with the formation of ferrons in the FeSi matrix at $T < T_C$ = 15 K. A comparison of the parameters $\chi_d(H, T)$ and $\chi_{loc}(H, T)$ shows that the estimated value of $\chi_{loc}(15\text{ K}) \approx 1.5 \times 10^{-5}$ cm^3/g (see Fig. 14) is almost an order of magnitude higher than the measured value of the volume-averaged susceptibility of FeSi, $\chi_d(15\text{ K}) \approx 2 \times 10^{-6}$ cm^3/g (see Fig. 7): $\chi_{loc}(15\text{ K})/\chi_d(15\text{ K}) \approx 8$. At the same time, to estimate the exchange enhancement factor reflecting the renormalization of the electron density of states $N^*(E_F)/N_0(E_F)$, it is obviously more appropriate to use the ratio of $\chi_{loc}(15\text{ K})$ to the value of magnetic susceptibility $\chi_d(T_{min} \approx 80\text{ K})$ corresponding to the transition to the regime of formation of a many-particle resonance in the vicinity of E_F [32]]. The exchange enhancement factor estimated in this way is

$$\frac{N^*(E_F)}{N_0(E_F)} \approx \frac{\chi_{loc}(15K)}{\chi_d(T_{min} \approx 80\text{ K})} \approx 20.$$

A close value of the exchange enhancement of the density of states, $N^*(E_F)/N_0(E_F) \geq 20$, was obtained in [22], where the Pauli paramagnetic contribution to $\chi_d(T)$ in FeSi was analyzed in the simple model of a narrow rectangular band ($\delta \approx 6$ meV) at the Fermi level. In our opinion, the estimates obtained for the local susceptibility substantiate the correctness of application of the spin-polaron approach to the description of the ground state of FeSi.

Concluding the analysis of the effect of positive magnetoresistance in FeSi, we consider the results of investigation of the magnetoresistance of FeSi in pulsed magnetic fields up to 400 kOe, which demonstrate a considerable positive magnetoresistance $\Delta\rho/\rho \approx 30\%$ (curve *4* in Fig. 4). In our opinion, this effect of positive magnetoresistance, which is observed for a low mobility of charge carriers ($\mu \leq 10$ cm^2/V s), can be explained by suppression of the many-particle resonance at the Fermi level, leading to the magnetic-field-induced metal–insulator transition under strong Hubbard correlations [45, 46]. In such a situation, the destruction of the spin-polaron resonance in a magnetic field must be accompanied by a considerable decrease in the Pauli susceptibility, which was indeed observed in the analysis of field dependences of parameters $\chi_d(H, T)$ and $\chi_{loc}(H, T)$ (see Figs. 7 and 14). Thus, the results of magnetoresistance measurements in a pulsed magnetic field may also serve as an additional confirmation of the applicability of the spin-polaron approach to the description of the ground state of FeSi.

Proceeding to analysis of the magnetoresistance of FeSi at temperatures $T < T_m \approx 7$ K, we emphasize that the sign inversion of magnetoresistance $\Delta\rho/\rho$ is observed in the immediate vicinity of the temperature corresponding to the transition to the mictomagnetic state, $T_m \approx 7$ K (see Fig. 4), and is accompanied by a change in the form of the field dependence of $\Delta\rho/\rho$ in weak magnetic fields (about 10 kOe) with a transition from the quadratic ($\Delta\rho/\rho \propto H^2$) to the linear ($\Delta\rho/\rho \propto H$) asymptotic behavior (see Figs. 2 and 3). The emergence of the linear field dependence of the magnetoresistance appears as rather nontrivial since the negative contribution to the magnetoresistance in magnetic materials in the general case is determined by the mutual orientation of the spins of charge carriers and scattering magnetic centers and is proportional to the square of magnetization; as a result, a negative magnetoresistance of the form $\Delta\rho/\rho \propto -M^2 \propto -H^2$ should be expected [47].

To explain the peculiar form of the field dependence of the negative magnetoresistance of FeSi at temperatures $T < T_m \approx 7$ K (see Fig. 3), we recall that the formation of the magnetic ground state in FeSi in the framework of the approach developed here is associated with the emergence of spontaneous polarization of spin–polaron surroundings; in this case, the local magnetization M of ferromagnetic regions is mainly determined by the exchange field at a charge carrier, $H_{ex} \approx 350$ kOe, and weakly depends on the external magnetic field in the limit $H \ll H_{ex}$. In such a situation, the total magnetization of the system can be represented as the sum of the field-independent spontaneous magnetization $M_0 = 4\pi H_{ex}$ of nanosize ferromagnetic regions and the additional contribution $\delta M(H) \propto \chi_{loc}$, which is determined by the polarization of a wide neighborhood of ferrons by the external magnetic field ($\delta M \ll M_0$). In this case, for the negative magnetoresistance of the paramagnetic matrix containing magnetic impurities, we will use the result obtained in [47], according to which

$$\rho(M) = \rho_0 - AM^2. \tag{6}$$

The, rewriting relation (6) in the form

$$\begin{aligned}\rho(M) &= \rho_0 - AM_0^2 - 2AM_0\delta M(H) - A(\delta M(H))^2 \\ &\approx \tilde{\rho}_0 - 2AM_0\delta M(H) - A(\delta M(H))^2\end{aligned} \tag{7}$$

and disregarding a small correction on the order of $(\delta M(H))^2$, we obtain

$$-\frac{\Delta\rho(H)}{\rho} \propto 2\delta M(H). \tag{8}$$

Thus, the variation of $\Delta\rho/\rho$ in a magnetic field is mainly determined by the field dependence $\delta M(H)$. Since the field dependence of the additional contribution $\delta M(H)$ to the magnetization for the system of ferromagnetic nanosize regions of small concentration can be approximated in the first approximation by the Brillouin function [39], for the negative magnetoresistance in FeSi we have

$$\begin{aligned}-\frac{\Delta\rho(H)}{\rho} &\propto \delta M(H) \propto B_J(\alpha) \\ &= \frac{2J+1}{2J}\coth\left(\frac{2J+1}{2J}\alpha\right) - \frac{1}{2J}\coth\left(\frac{1}{2J}\alpha\right),\end{aligned} \tag{9}$$

where $\alpha = \mu_{eff}H/k_BT$, J is the total magnetic moment, and $\mu_{eff} = g\mu_B J$ is the effective magnetic moment of ferromagnetic nanoclusters. As we proceed to the classical limit $J \longrightarrow \infty$, the Brillouin function $B_J(\alpha)$ is transformed into the Langevin function $L(\alpha)$; as a result, the negative magnetoresistance in FeSi is defined as

$$-\frac{\Delta\rho(H)}{\rho} \propto L(\alpha) = \coth\alpha - 1/\alpha. \tag{10}$$

In the limit of a weak magnetic field $\alpha \ll 1$, expression (9) (as well as its classical analogue (10)) obviously leads to a linear field dependence $-d\rho/\rho \propto \alpha \propto H$ for the negative magnetoresistance.

Figure 15 presents the results of quantitative analysis of the negative magnetoresistance in FeSi. To eliminate the additional contribution from the positive magnetoresistance, which is observed in the immediate

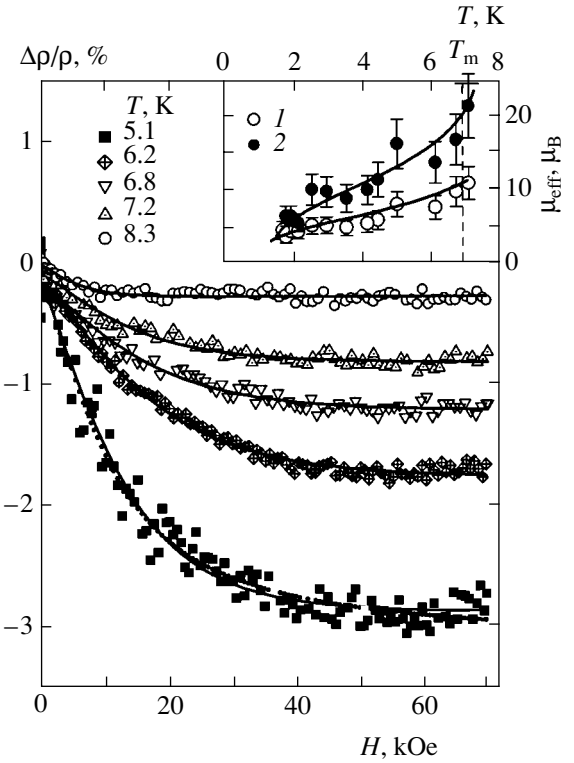


Fig. 15. Field dependence of the negative contribution to the magnetoresistance $\Delta\rho/\rho = (\Delta\rho/\rho)_{exp} - AH^2$ in the temperature range above the helium temperature. Solid curves show the results of approximation of the experimental data by the Brillouin function $B_J(\alpha)$ with $J = 1$ (formula (8)). The dotted curve shows the result of simulation of the Langevin function $L(\alpha)$ for the field dependence of $\Delta\rho/\rho$ measured at $T = 5.1$ K (formula (9)). The inset shows the temperature dependences of the effective magnetic moment $\mu_{eff}(T)$ estimated in the framework of model (6)–(9) at $T < T_m \approx 7$ K for $B_J(\alpha)$ (*1*) and $L(\alpha)$ (*2*).

vicinity of the transition to the mictomagnetic state, $T \sim T_m \approx 7$ K (see the curves of $\Delta\rho/\rho$ in the temperature range 6–8 K in Fig. 3a), we subtracted from the experimental dependences of $\Delta\rho/\rho$ the quadratic asymptotic form $\Delta\rho/\rho \propto H^2$ in magnetic fields stronger than 40 kOe. The results presented in Fig. 15 show that approximation (6)–(10) makes it possible to describe the field dependence of the negative magnetoresistance to a high degree of accuracy in the entire range of magnetic fields used in the present study. It should also be noted that a comparison of the results of approximation of the $\Delta\rho/\rho$ curves in the quantum ($J = 1$) (9) and classical (10) cases does not permit an unambiguous choice between these two cases (see the $\Delta\rho/\rho$ dependence simulated in Fig. 15 for $T = 5.1$ K). At the same time, the temperature dependences of the effective magnetic moment μ_{eff} estimated from the results of numerical fitting of the parameters given by expressions (9) and (10) (see the inset to Fig. 15) demonstrate a decrease in the effective magnetic moment with decreasing temperature from $\mu_{eff} \approx 12\mu_B$ ($20\mu_B$) at $T \sim T_m \approx 7$ K to $\mu_{eff} \approx 4\mu_B$ ($7\mu_B$) at $T = 1.8$ K (the values of μ_{eff} for calculations using expression (10) are given in parentheses). The observed qualitative agreement in the behavior of the effective magnetic moment μ_{eff} calculated using relations (9) and (10) indicates a substantial decrease in the effective magnetic moment, which is obviously due to the "actuation" of the interaction between nanosize ferromagnetic regions and, as a consequence, due to "freezing" of the magnetic moments of nanoclusters during the formation of the mictomagnetic ground state in the FeSi matrix at $T < T_m \approx 7$ K (SG phase in Fig. 1c).

4.3. Low-Temperature Magnetic Phase Diagram and Parameters of Charge Carriers in FeSi

To construct the low-temperature magnetic phase diagram of FeSi, we generalize the results of the present analysis of transport, magnetic, and magnetooptical characteristics of this compound as well as the experimental results known from the literature [22, 23, 32].

The temperature range 15 K < T < 100 K corresponds to the paramagnetic phase (region P in Fig. 16) of FeSi, in which strong Hubbard correlations lead to a substantial renormalization of the density of states, accompanied by the formation of a many-particle resonance in the vicinity of the Fermi energy (structure I in Fig. 1), which determines physical properties of FeSi at low temperatures. Analysis of transport characteristics [22, 23] shows that many-particle states within the resonance should be juxtaposed to a system of spin polarons having a small radius $r_{SP} \approx 5$ Å and a concentration $n \sim 10^{17}$–10^{18} cm^{-3} with the characteristic binding energy $E_p \approx 6$ meV and with a noticeably increased effective mass $m^* \sim 100m_0$. The substantial increase in the Pauli susceptibility observed upon a decrease in temperature in the interval 15 K < T < 100 K [32] enables us to directly estimate the increase in the electron density of states at the Fermi level due to the emergence of the many-particle resonance for $N \geq 20$ states/(eV unit cell).

It has been established in this study that a decrease in temperature in the vicinity of $T_C = 15$ K in FeSi is accompanied by anomalous Hall effect $R_H^A(T)$ corresponding to the electron type of conduction, $R_H^A < 0$ (see Figs. 9 and 12a). Taking into account the increase in the local susceptibility χ_{loc} observed near T_C (see the inset to Fig. 14) and in the local magnetization M_{loc} (see Fig. 13), the emergence of anomalies in the transport characteristics of FeSi in the temperature range 7 K < T < 15 K should be attributed to the emergence of the coherent regime of spin fluctuations and to the transformation of spin polarons into nanosize ferromagnetic regions (ferrons) (phase F in Fig. 1b and in Fig. 16). This conclusion is confirmed by the emergence of hysteresis in the angular dependences of the Hall resistance in a magnetic field $H < 3.6$ kOe at temperatures below

T_C = 15 K [23] (symbols *1* in Fig. 16). Our analysis of the angular dependences of the Hall resistance shows that the ferrons formed in the above-mentioned temperature range are characterized by uniaxial magnetic anisotropy with a field of $H_{an} = 3 \pm 1$ kOe (symbols *2* and phase F in Fig. 16). It should be emphasized that the emergence of spontaneous magnetization in FeSi at $T < 15$ K correlates with the fulfillment of the ferromagnetism criterion similar to the Stoner criterion ($UN > 1$), which has the form $UN \geq 5$ for a Hubbard correlation energy of $U \approx 270$ meV in FeSi [23]. At the same time, the low concentration of ferrons (10^{17}–10^{18} cm^{-3}) is responsible only for weak singularities of the "integrated" magnetic characteristics (symbols *3* in Fig. 16) since the magnetization and susceptibility of FeSi in this temperature range are mainly determined by the paramagnetic contribution of the FeSi matrix [23, 32].

Additional estimates of the charge carrier parameters in the ferromagnetic (F) state of FeSi can be obtained from the temperature dependences of effective charge carrier concentrations (see Fig. 12a). Under the conditions of a strong electron–phonon interaction, the linewidth of optical phonons in the frequency range 180–400 cm^{-1}, γ_i^{ph} (198 cm^{-1}, 318 cm^{-1}, 338 cm^{-1}) ≈ 10 cm^{-1} [48], can be used for estimating the relaxation time for charge carriers in FeSi:

$$\langle \tau_{e-ph} \rangle = \frac{1}{3\gamma_i^{ph}} \approx 1.7 \times 10^{-13} \text{ s}.$$

Analogous values of $\langle \tau_{e-ph} \rangle$ can be obtained from the reciprocal linewidth of magnetic scattering of polarized neutrons Γ = 3–4 meV [49]. Using the expression for the Hall mobility $\mu = R_H/\rho = e\tau/m^*$, we can estimate the effective mass of charge carriers in FeSi. The obtained values of m^*/m_0 calculated for magnetic fields of 17.2 and 70.3 kOe are shown in Fig. 12b. It should be noted that the values $m^*/m_0 = 80 \pm 20$ corresponding to many-particle states in the FeSi matrix at $T > 10$ K (see Fig. 12b) correlate well (to within the experimental error) with the estimate $m^*/m_0 \approx 50$ obtained from the results of measurements of optical conductivity $\sigma(\omega)$ in the frequency range 50–25000 cm^{-1} [50]. The coincidence of the values of effective mass of charge carriers—spin polarons in the paramagnetic phase ($T > T_C \approx$ 15 K) and ferrons in the ferromagnetic phase ($T < T_C \approx$ 15 K) (see Fig. 12b)—can be regarded as an additional argument in favor of the proposed many-particle description of the low-temperature transport and genesis of anomalies in the Hall effect in FeSi.

Analyzing the characteristics of the ferromagnetic phase (F in Fig. 16) at $T < T_C \approx 15$ K, we must take into account the results of measurements of the low-temperature heat capacity $C(T)$ of monocrystalline FeSi samples [14], according to which a broad peak is present on the $C(T)$ curves in the vicinity of temperature $T \approx 8.5$ K.

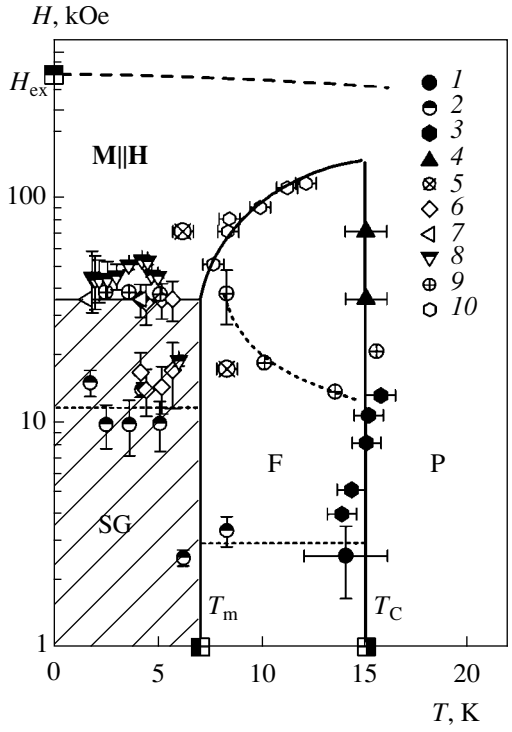


Fig. 16. Low-temperature magnetic phase diagram of FeSi. The symbols used in the diagram correspond to the emergence of hysteresis ΔR_{hyst} on the angular dependences of the Hall resistance [23] (*1*); the temperature dependence of anisotropy field H_{an} of nanosize ferromagnetic regions (*2*); peculiarities in the temperature dependences of magnetization $M(T)$, measured in fields up to 12.5 kOe [23] (*3*); the peak on the temperature dependence of local susceptibility χ_{loc} (see text and Fig. 14) (*4*); the peculiarity on the temperature dependence of the spin-polar component R_H^{SP} of the Hall effect (Fig. 12a) (*5*); the anomalies (hysteresis) of the microwave magnetooptical absorption in the vicinity of fields $H_{an} \approx 15$ kOe and $H_m \approx 35$ kOe (Fig. 8) (*6*); the magnetic field determining the emergence of the anisotropic contribution to magnetoresistance $\Delta\rho/\rho$ [23] (*7*); the magnetic field corresponding to the point of inflection H_s on the field dependences of magnetoresistance $\Delta\rho/\rho$ (Fig. 3b) (*8*); the change in the asymptotic behavior of the second harmonic ρ_{H2} of the Hall resistance (Fig. 10) (*9*), and the peak of differential susceptibility $\chi_d(T)$ (Fig. 7) (*10*). Phase notation is the same as in Fig. 1.

Analysis of this singularity in the framework of a simple two-level model (Schottky anomaly) makes it possible to estimate the characteristic energy of splitting of the narrow band of many-particle states in the vicinity of E_F: $\Delta = 2$ meV [14]. Assuming that the emergence of ferromagnetism at temperatures $T < T_C \approx 15$ K under the conditions of the strong ($UN \geq 5$) interaction is accompanied by exchange-induced splitting of the many-particle resonance of width $\delta \approx 6$ meV at the Fermi level (structure II in Fig. 1), we can directly estimate the exchange field at a carrier in FeSi from the results of thermal measurements [14]: $H_{ex} = \Delta/\mu_B \approx$ 350 kOe. In our opinion, the good coincidence of the

obtained result and the value of $H_{ex} \approx 350 \pm 100$ kOe obtained from simultaneous analysis of the transport and magnetic parameters of FeSi [23, 32] is a sound argument in favor of the spin-polaron approach used here for describing the low-temperature properties of FeSi.

Passing to the region of $T < T_m = 7$ K, $H < H_m \approx 35$ kOe on the H–T phase diagram of FeSi, we note that anomaly of the spin-polaron component (R_H^{SP}) of the Hall effect is observed upon a decrease in temperature in the vicinity of $T_m = 7$ K in a weak magnetic field (see Fig. 12a and symbols *5* in Fig. 16). In addition at $T_m \leq 7$ K, the dependences $\Delta\rho/\rho = f(T, H)$ acquire a negative contribution to the magnetoresistance (see curve *3* in Fig. 4). The effects listed above apparently reflect the renormalizaion of parameters of ferromagnetic regions. The considerable increase in the anisotropy field of nanoclusters to values of $H_{an} \approx 12 \pm 2$ kOe (symbols *2* and region SG in Fig 16) and the sharp change in the phase of the second harmonic of the Hall signal (see the inset to Fig. 10) also speak in favor of such an interpretation. In addition, our analysis of galvanomagnetic data shows that the transition at $T < T_m = 7$ K is accompanied by a considerable decrease in the Hall concentration and the effective mass of charge carriers to values of $p \sim 10^{16}$–10^{17} cm^{-3} and $m^* \approx (20 \pm 5)m_0$, respectively (see Fig. 12). Taking into account the substantial decrease in the effective magnetic moment μ_{eff} of ferromagnetic regions from $(10–20)\mu_B$ at $T \sim T_m \approx 7$ K to $(4–6)\mu_B$ at $T = 1.8$ K (see the inset to Fig. 15), we conclude that the observed changes in the above parameters obviously indicate the formation of ferromagnetic microregions from interacting ferrons upon a transition to the mictomagnetic state at $T \approx T_m = 7$ K (phase SG in Fig. 1c and Fig. 16). In our opinion, the anomalies on the temperature dependences of the above parameters near $T_m \approx 7$ K (see Fig. 12) should be attributed to the enhancement of scattering by inhomogeneities in the immediate vicinity of the transition to the mictomagnetic state and, as a consequence, to a sharp decrease in the mobility of charge carriers.

Comparing FeSi with a classical band ferromagnetic, it should be emphasized that the behavior of physical parameters of FeSi is determined by the direct participation of nanosize magnetic regions in the charge-transfer processes at low temperatures. In such a situation, along with the emergence of anomalies in the physical properties typical of mictomagnetic systems in FeSi, the possibility of formation of new spatially organized magnetic structures, the realization of magnetic orientational transitions, etc., cannot be ruled out either. In our opinion, it is the magnetic transition in the structure of nanosize ferromagnetic regions that is responsible for the peculiarities in the physical parameters of FeSi observed in a magnetic field of $H_m \approx 35$ kOe at $T < T_m \approx 7$ K. The most noticeable anomalies (hysteresis) in these ranges of temperatures and magnetic fields are observed in the field dependences of magnetooptical absorption (see Fig. 8 and symbols *6* in Fig. 16). In addition, in a magnetic field of $H \approx 35$ kOe, an anisotropic contribution to the magnetoresistance is observed [23] (symbols *7* in Fig. 16) and the $\Delta\rho/\rho$ curves acquire a point of inflection H_s(symbols *8* in Fig. 16). For $H \leq 35$ kOe, the asymptotic behavior of the field dependences of the second harmonic amplitude in the Hall resistance also changes (see Fig. 10 and symbols *9* in Fig. 16). In our opinion, these peculiarities indicate the change in the magnetic structure of FeSi, which is accompanied by the rupture of bonds between ferromagnetic nanoclusters (phase SG in Fig. 1c), followed by the reorientation of the magnetic moments of ferrons parallel to the applied magnetic field (region **M** || **H** in Fig. 1d and in Fig. 16).

An additional argument in favor of realization of the magnetic orientational transition at $T < T_m \approx 7$ K accompanied by the destruction of the mictomagnetic ground state of FeSi in magnetic fields above $H_m \approx 35$ kOe can be the increase in the effective mass of charge carriers to values of $m^* \approx (70 \pm 20)m_0$ in the field $H = 70.3$ kOe (see Fig. 12b). The observed "increase" in the mass of charge carriers in FeSi with increasing m^* to values of the effective mass of spin polarons in the paramagnetic (P) phase and ferrons in the ferromagnetic (F) phase can be naturally attributed to the destruction of intracluster bonds in the mictomagnetic (SG) phase by an external electric field. The polarization of magnetic moments of nanosize ferromagnetic regions induced by an external field leads to a transition to a state with ferrons oriented along the magnetic field upon an increase in H in the vicinity of $H_m \approx 35$ kOe (state **M** || **H** in Fig. 1d). In addition, the peak T_{max}^{χ} observed on the temperature dependences of the differential susceptibility during the measurement of magnetic characteristics (see Fig. 7 and symbols *10* in Fig. 16) as well as the singularity on the temperature dependence of the spin-polaron component R_H^{SP} of the Hall effect in a magnetic field $H > H_m$ (see Fig. 12a and symbols *5* in Fig. 16) should apparently be also interpreted in terms of the phase boundary and, accordingly, the new phase **M** || **H** on the low-temperature phase diagram of FeSi. Additional experiments including the measurement of transport and magnetic characteristics of FeSi in the range of magnetic fields corresponding to the values of the exchange field of spin-polaron states are required to determine the structure of the low-temperature phase diagram and, in particular, the peculiarities of the phase transition between the collinear magnetic (**M** || **H**) and paramagnetic (P) phase in FeSi.

5. CONCLUSIONS

The above analysis of galvanomagnetic characteristics of FeSi has made it possible for the first time to study in detail the features of low-temperature transport

and to determine the characteristics of charge carriers in wide ranges of temperatures (1.8–40 K) and magnetic fields (up to 120 kOe), corresponding to the formation of a magnetic state in this narrowband semiconductor with strong quasiparticle interactions. It is shown that low-temperature anomalies of the Hall resistance in FeSi, including the sign inversion in the Hall coefficient at $T \approx 12$–15 K, are caused by a complex competition between the spin-polaron ($\rho_H^{SP} > 0$) and the anomalous magnetic ($\rho_H^A < 0$) contributions to the Hall coefficient. At $T < 10$ K, the Hall signal is determined by the anomalous component ρ_H^A of the Hall resistance, whose absolute value increases by more than five orders of magnitude in the temperature interval 1.8–20 K and attains values of about 1 Ω cm at $T = 1.8$ K.

The above analysis of experimental data shows that the anomalous Hall effect in FeSi can apparently be associated with stabilization of the coherent mode of spin fluctuations in the formation of nanosize ferromagnetic regions (ferrons) (of approximately 10 Å in diameter) near $T_C = 15$ K. For a quantitative description of the Hall effect in FeSi, we propose a procedure for separating the Hall resistance components, in which the set of experimental data is represented in terms of the "normal" (spin-polaron) R_H^{SP} and anomalous R_H^A Hall coefficients, with a subsequent estimation of the effective parameters of charge carriers in low-temperature magnetic phases of FeSi.

Analysis of the angular dependences of the Hall resistance ρ_H in FeSi revealed the presence of the additional harmonic $\rho_{H2}(\varphi) \propto \cos 2\varphi$ leading to substantial deviations of the experimental curves from the standard cosinusoidal signal. It is found that this second harmonic, which dominates in the immediate vicinity of the temperature corresponding to the sign inversion in the Hall coefficient at $T \approx 12$–15 K is associated with the emergence with a magnetic-field-even contribution to the Hall effect that cannot be explained in the framework of traditional phenomenological models.

Analysis of the temperature dependences of the magnetoresistance of FeSi enabled us to estimate some characteristics of many-particle states realized under rapid spin fluctuations at Fe centers. It was shown that the positive magnetic contribution to the magnetoresistance $\Delta\rho/\rho$ observed in the vicinity of the Curie temperature $T_C = 15$ K emerges as a result of polarization of the local surroundings of a charge carrier upon a transition to the coherent mode of spin fluctuations in FeSi. It is interesting to note that the sign inversion of $\Delta\rho/\rho$ observed upon a decrease in temperature in the immediate vicinity of the mictomagnetic phase boundary at $T_m = 7$ K is accompanied by a transition from a positive ($\Delta\rho/\rho > 0$, $T > T_m$) to a negative ($\Delta\rho/\rho < 0$, $T < T_m$) magnetoresistance. The model proposed here for describing the anomalous linear asymptotic behavior of $\Delta\rho/\rho \propto -H$, which is observed in weak magnetic fields (around 10 kOe) in the negative magnetoresistance mode, makes it possible to associate the observed effect with peculiarities in the formation of a magnetic state in a system of interacting ferrons in the mictomagnetic phase of FeSi.

The analysis of the field and temperature dependences of transport characteristics enables us to obtain quantitative estimates for the effects of exchange enhancement as well as microscopic parameters of many-particle complexes determining the regimes of charge transfer under rapid (on the order of 10^{-13} s) electron density fluctuations at Fe centers. On the basis of the results obtained in this study and indicating the realization of a sequence of transformations of many-particle states (spin polarons ⟶ ferrons ⟶ magnetic nanoclusters) during the formation of the ferromagnetic and mictomagnetic phases of FeSi, we constructed for the first time the low-temperature magnetic phase diagram of this compound. The aggregate of anomalies in the transport and magnetic characteristics discovered in the vicinity of $H_m \approx 35$ kOe in the mictomagnetic state of FeSi at $T \leq T_m = 7$ K leads to the conclusion concerning the magnetic-field-induced transition to a new collinear phase $\mathbf{M} \parallel \mathbf{H}$ on the phase diagram of FeSi. At the same time, additional investigations of the transport and magnetic parameters of FeSi with strong electron correlations are required to clarify the structure and parameters of the low-temperature state of this compound in the range of magnetic fields corresponding to the exchange fields $H_{ex} \approx 350$ kOe of spin-polaron states.

ACKNOWLEDGMENTS

This study was financed by the Russian Foundation for Basic research (project nos. 01-02-16601 and 03-02-06531), the Science and Engineering Program "New Materials" of the Ministry of Education of the Russian Federation (project no. 202.07.01.023), INTAS (grant no. 03-51-3036), the program "Strongly Correlated Electrons in Semiconductors, Metals, Superconductors, and Magnetic Materials" of the Physics Division of the Russian Academy of Sciences, the Program for the Development of an Instrumental Basis of Scientific Institutions of the Ministry of Industry and Science of the Russian Federation, and the Program for Support of Young Scientists of the Russian Academy of Sciences. Two of the authors (V. V. G. and S. V. D) are grateful for individual support from the Foundation for Promoting Russian Science.

REFERENCES

1. T. Moriya, Usp. Fiz. Nauk **135**, 117 (1981).
2. T. Moriya, *Spin Fluctuations in Itinerant Electron Magnetism* (Springer, Berlin, 1985; Mir, Moscow, 1988).

3. S. N. Evangelou and D. M. Edwards, J. Phys. C **16**, 2121 (1983).
4. G. Shirane, J. E. Fisher, Y. Endoh, *et al.*, Phys. Rev. Lett. **59**, 351 (1987).
5. V. I. Anisimov, S. Yu. Ezhov, I. S. Efimov, *et al.*, Phys. Rev. Lett. **76**, 1735 (1996).
6. E. Kulatov and H. Ohta, J. Phys. Soc. Jpn. **66**, 2386 (1997).
7. E. Kulatov, H. Ohta, T. Arioka, *et al.*, J. Phys.: Condens. Matter **9**, 9043 (1997).
8. H. Yamada, K. Terao, H. Ohta, *et al.*, J. Phys.: Condens. Matter **11**, L309 (1999).
9. O. M. Tatsenko and V. D. Selemir, in *Abstracts of 8th International Conference on Megagauss Magnetic Field Generation and Related Topics* (Tallahassee, USA, 1998).
10. Yu. B. Kudasov, A. G. Volkov, A. A. Povzner, *et al.*, Zh. Éksp. Teor. Fiz. **116**, 1770 (1996) [JETP **89**, 960 (1996)].
11. J. F. DiTusa, K. Friemelt, E. Bucher, *et al.*, Phys. Rev. Lett. **78**, 2831 (1997).
12. J. F. DiTusa, K. Friemelt, E. Bucher, *et al.*, Phys. Rev. B **58**, 10288 (1998).
13. G. Föex, J. Phys. Radium **9**, 37 (1938).
14. M. A. Chernikov, L. DeGiorgi, E. Felder, *et al.*, Phys. Rev. B **56**, 1366 (1997).
15. S. Pashen, E. Felder, M. A. Chernikov, *et al.*, Phys. Rev. B **56**, 12916 (1997).
16. M. Fäth, J. Aarts, A. A. Menovsky, *et al.*, Phys. Rev. B **58**, 15483 (1998).
17. Y. Takahashi, J. Phys.: Condens. Matter **10**, L671 (1998).
18. K. Urasaki and T. Saso, Phys. Rev. B **58**, 15528 (1998).
19. D. van der Marel, A. Damascelli, K. Schulte, *et al.*, Physica B (Amsterdam) **244**, 138 (1998).
20. P. S. Riseborough, Phys. Rev. B **58**, 15534 (1998).
21. T. Jarlborg, Physica B (Amsterdam) **293**, 224 (2001).
22. N. E. Sluchanko, V. V. Glushkov, S. V. Demishev, *et al.*, Europhys. Lett. **51**, 557 (2000).
23. N. E. Sluchanko, V. V. Glushkov, S. V. Demishev, *et al.*, Zh. Éksp. Teor. Fiz. **119**, 359 (2001) [JETP **92**, 312 (2001)].
24. A. Georges, G. Kotliar, W. Krauth, *et al.*, Rev. Mod. Phys. **68**, 13 (1996).
25. N. E. Sluchanko, V. V. Glushkov, S. V. Demishev, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **68**, 774 (1998) [JETP Lett. **68**, 817 (1998)].
26. N. E. Sluchanko, A. V. Bogach, V. V. Glushkov, *et al.*, Zh. Éksp. Teor. Fiz. **125**, 906 (2004) [JETP **98**, 793 (2004)].
27. B. C. Sales, E. C. Jones, B. C. Chakoumakos, *et al.*, Phys. Rev. B **50**, 8207 (1994).
28. N. E. Sluchanko, A. V. Bogach, I. B. Voskoboĭnikov, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **45**, 1046 (2003) [Phys. Solid State **45**, 1096 (2003)].
29. N. E. Sluchanko, S. V. Demishev, A. V. Semeno, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **63**, 431 (1996) [JETP Lett. **63**, 453 (1996)].
30. N. E. Sluchanko, V. V. Glushkov, S. V. Demishev, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 745 (1999) [JETP Lett. **69**, 798 (1999)].
31. E. V. Kuchis, *Galvanomagnetic Effects and Methods of Their Studies* (Radio i Svyaz', Moscow, 1990), p. 115 [in Russian].
32. N. E. Sluchanko, V. V. Glushkov, S. V. Demishev, *et al.*, Phys. Rev. B **65**, 064404 (2002).
33. N. E. Sluchanko, A. V. Bogach, V. V. Glushkov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 31 (2002) [JETP Lett. **76**, 26 (2002)].
34. É. L. Nagaev and É. B. Sokolova, Fiz. Tverd. Tela (Leningrad) **19**, 732 (1977) [Sov. Phys. Solid State **19**, 425 (1977)].
35. S. H. Chun, M. B. Salamon, Y. Tomioka, *et al.*, Phys. Rev. B **61**, R9225 (2000).
36. Y. Lyanda-Geller, S. H. Chun, M. B. Salamon, *et al.*, Phys. Rev. B **63**, 184426 (2001).
37. Jinwu Ye, Yong Baek Kim, A. J. Millis, *et al.*, Phys. Rev. Lett. **83**, 3737 (1999).
38. R. Resta, J. Phys.: Condens. Matter **12**, R107 (2000).
39. G. S. Krinchik, *Physics of Magnetic Phenomena* (Mosk. Gos. Univ., Moscow, 1985) [in Russian].
40. P. M. Chaikin, in *Organic Superconductivity*, Ed. by V. Z. Kresin and W. A. Little (Plenum, New York, 1990).
41. M. B. Hunt, M. A. Chernikov, E. Felder, *et al.*, Phys. Rev. B **50**, 14933 (1994).
42. P. Coleman, P. W. Anderson, and T. V. Ramakrishnan, Phys. Rev. Lett. **55**, 414 (1985).
43. A. Fert and P. M. Levy, Phys. Rev. B **36**, 1907 (1987).
44. S. Chikasumi, *The Physics of Ferromagnetism. Magnetic Characteristics and Engineering Applications* (Syokabo, Tokyo, 1980; Mir, Moscow, 1983), Vol. 1.
45. L. Laloux, A. Georges, and W. Krauth, Phys. Rev. B **50**, 3092 (1994).
46. Ph. Nozieres, Eur. Phys. J. B **6**, 447 (1998).
47. K. Yosida, Phys. Rev. **107**, 396 (1957).
48. A. Damascelli, K. Schulte, D. Van der Marel, *et al.*, Phys. Rev. B **55**, R4863 (1997).
49. K. Tajima, Y. Endoh, J. E. Fisher, *et al.*, Phys. Rev. B **38**, 6954 (1988).
50. Z. Schlesinger, Z. Fisk, Hai-Tao Zhang, *et al.*, Phys. Rev. Lett. **71**, 1748 (1993).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 99, No. 2, 2004, pp. 415–416.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 2, 2004, pp. 469–471.

SOLIDS
Electronic Properties

Self-Induced Supertransparency in Small-Size Injection Lasers

V. P. Strakhov

Lebedev Physical Institute, Russian Academy of Sciences, Moscow, 119991 Russia

Received December 9, 2003

Abstract—The emission spectrum of an injection GaAs laser having a quadrilateral cavity with a 13×13 μm square cross section and with a high pumping level is considered. Anomalously high dispersion revealed for the refractive index for the short-wave mode ($\tilde{n} > 7$) and disappearance of dispersion for the long-wave ($\tilde{n} = 3.4$) mode indicate in all probability the emergence of a gap in the electron energy spectrum at the wavelength of the fundamental lasing mode. In contrast to [1], the entire long-wave part of the spectrum is emitted coherently at the same frequency and probably without absorption. © *2004 MAIK "Nauka/Interperiodica".*

We analyze here injection GaAs lasers with a quadrilateral cavity with a 13×13 μm square cross section, obtained by chipping along the cleavage planes from the initial 80-μm-thick plate with a *p–n* junction, which was grown by liquid-phase epitaxy. For better visualization, we can imagine that about 30 such lasers can be accommodated on a cross section of human hair. This is the smallest known laser in the world.

It is important to note that the optical length of such a resonator is $L < 37$ μm, and the photon transit time in such a resonator is $t < 5 \times 10^{-13}$ s, which is shorter than the characteristic thermal relaxation time T_2 in the semiconductor at 77 K. This makes it possible to learn about absolutely new features of operation of injection lasers.

Figure 1a shows the emission spectrum of such a laser at the lasing threshold $I_{th} = 0.7$ mA. Lasing begins at the threshold simultaneously for all equidistant axial modes on the entire long-wave part of the laser amplification band. The separation $\Delta\lambda \approx 21$ Å between the modes corresponds to the resonator length $L = 37$ μm and is defined by the relation

$$\Delta\lambda = \frac{\lambda_1\lambda_2}{2L\tilde{n}}, \qquad (1)$$

where

$$\tilde{n} = n_0 - \lambda\frac{dn}{d\lambda}$$

is the effective refractive index taking dispersion into account (n_0 is the refractive index of pure material). Estimates of the spectrum give the value $\tilde{n} = 5.156$, which matches the results obtained by other authors for various materials [2–4].

It was mentioned in the previous publication [1] that the effect is preserved for intensities two or three times higher than the threshold I_{th}, after which the breakdown occurs as a rule. Neither the thermal nor the optical breakdown mechanism operates in this case. In recent years, laser-based optoacoustics is being actively developed. Photoexcitation of hypersound pulses in semiconductors and waveguides is studied extensively by

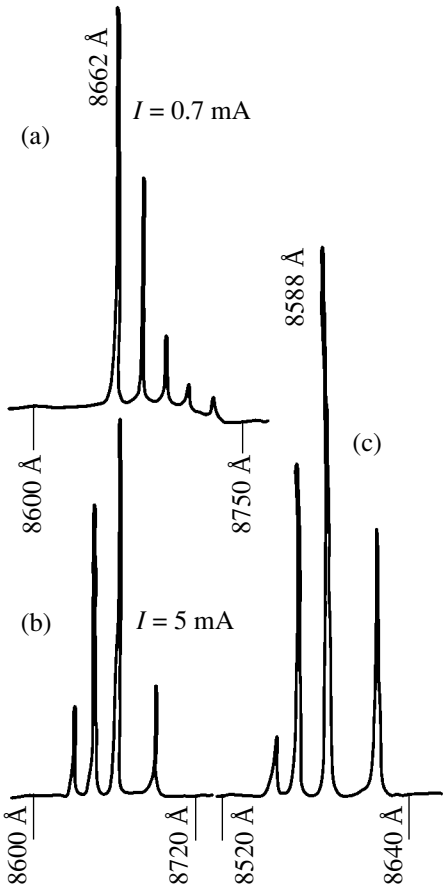


Fig. 1.

various groups of researchers [5–8]. Excess voltage was detected by the author more than three decades ago [9, 10] in powerful injection-type cw parametric quantum oscillators [11]. The possibility of hypersound generation leading to a gradual degradation of such lasers at high pumping densities was considered. The difficulties encountered in assembling ultrasmall lasers are apparent. We used a holder with a spring-loaded contact. It is very difficult to arrange the crystal strictly coaxially. For an area $S \approx 10^{-6}$ cm^2, a force of 1 gf leads to pressures of 1000 atm on the crystal. For this reason, acoustic perturbations emerging during laser operation above the threshold just expel the crystal from the spring-loaded holder or crush it.

However, the most successfully assembled crystals operated at intensities an order of magnitude higher than I_{th}.

The spectrum of such a laser with an intensity seven times higher than I_{th} is shown in Fig. 1b. The predominant mode $\lambda = 8662$ Å is present in the spectra for all currents, indicating the absence of heating. At the same time, the spectra are obviously quite different.

It is hardly appropriate to speak here about uniform broadening of the amplification band. It was shown in [12] that if electron transition between energy bands occur over time periods shorter than the time of interaction with phonons, the kinetic equations for electrons should be replaced by the kinetic equations for quasiparticles. The physical meaning corresponding to this diagram is that direct transitions are saturated and the processes of emission and absorption of electromagnetic field quanta occur due to indirect transitions accompanied with the emission of acoustic phonons.

In the field of an intense electromagnetic wave [13, 14], this leads to the emergence of a new steady state in which the wave absorption substantially decreases on account of reverse transitions of electrons under the action of the field.

Using Fig. 1b and formula (1), we find that the following expression holds for the central mode:

$$\tilde{n} = n_0 - \lambda\frac{dn}{d\lambda} \approx 5.15$$

(the error in the measurements of the line is about 5%). For the short-wave mode, the effective refractive index abruptly increases to $\tilde{n} > 7$, indicating the intensification of dispersion and absorption.

For the long-wave mode, $\tilde{n} \approx 3.4$; i.e., the dispersion in the refractive index disappears completely,

$$\lambda\frac{dn}{d\lambda} = 0,$$

and radiation propagates without losses. In this case, the entire long-wave radiation in Fig. 1a is strictly phased and coherent and is emitted at the same frequency.

Such an anomalous behavior of dispersion in all probability indicates the formation of a gap in the electron energy spectrum at the wavelength of the dominant mode.

Figure 1c shows the spectrum of a similar laser made of another weakly doped material. The spectrum is shifted by almost 100 Å to the short-wave part. It can be seen that the spectrum is completely identical to that in Fig. 1b: it contains the same mode and exhibits the same behavior of dispersion. This indicates the fundamental behavior of such lasers.

ACKNOWLEDGMENTS

The author is grateful to A.F. Andreev for his interest in his research as well as to Yu.V. Kopaev and V.F. Elesin for fruitful discussions.

REFERENCES

1. V. P. Strakhov, Zh. Éksp. Teor. Fiz. **123**, 1276 (2003) [JETP **96**, 1122 (2003)].
2. H. Kogelnik and S. V. Shank, J. Appl. Phys. **43**, 2327 (1972).
3. J. Zoroofchi and J. K. Butler, J. Appl. Phys. **44**, 3697 (1973).
4. M. Nakamura *et al.*, IEEE J. Quantum Electron. **11**, 436 (1975).
5. V. É. Gusev and A. A. Karabutov, *Laser Optical Acoustics* (Nauka, Moscow, 1991) [in Russian].
6. N. V. Chigarev, D. Yu. Parashchuk, Yu. S. Pan, and V. É. Gusev, Zh. Éksp. Teor. Fiz. **121**, 728 (2002) [JETP **94**, 627 (2002)].
7. N. V. Chigarev, Kvantovaya Élektron. (Moscow) **32**, 844 (2002).
8. A. E. Biryukov, M. E. Sukharev, and E. M. Dianov, Kvantovaya Élektron. (Moscow) **32**, 765 (2002).
9. V. P. Strakhov, in *Abstracts of IV All-Union Conference on Nonlinear Acoustics* (Ashkhabad, 1969).
10. V. P. Strakhov, in *Abstracts of V All-Union Conference on Nonlinear Optics* (Kishinev, 1970).
11. P. G. Eliseev and V. P. Strakhov, Zh. Tekh. Fiz. **40**, 1564 (1970) [Sov. Phys. Tech. Phys. **15**, 1214 (1970)].
12. V. M. Galitskiĭ and V. F. Elesin, Zh. Éksp. Teor. Fiz. **68**, 216 (1975) [Sov. Phys. JETP **41**, 104 (1975)].
13. V. F. Elesin, Fiz. Tverd. Tela (Leningrad) **11**, 1820 (1969) [Sov. Phys. Solid State **11**, 1470 (1970)].
14. I. A. Poluéktov, Doctoral Dissertation in Physics and Mathematics (Physical Inst., USSR Academy of Sciences, Moscow, 1981).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 99, No. 2, 2004, pp. 417–426.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 2, 2004, pp. 472–482.

SOLIDS
Electronic Properties

Modulated Magnetic Structure of Weak Rhombohedral Ferromagnets α-Fe_2O_3:Ga and $FeBO_3$:Mg

B. Yu. Sokolov

Ulugbek National University of Uzbekistan, Vuzgorodok, Tashkent, Republic of Uzbekistan
e-mail: optic@nuuz.uzsci.net

Received January 6, 2004

Abstract—The effect of diamagnetic impurities on the stability of the homogeneous magnetic state of rhombohedral antiferromagnets with weak ferromagnetism (α-Fe_2O_3:Ga and $FeBO_3$:Mg) is studied experimentally. It is shown that the application of an external magnetic field in the basal plane in the crystals under study in a certain temperature range induces a magnetic superstructure along the hard magnetization axis, which can be presented in the form of a ripplon phase with the azimuth of the local ferromagnetism vector oscillating about the direction of the field. The preferred orientation of the discovered modulated structures relative to crystallographic directions in the basal plane of α-Fe_2O_3:Ga and $FeBO_3$:Mg is studied, and the dependence of the spatial period of the superstructure on the applied magnetic field and temperature is analyzed. The magnetic-field-induced transition of the studied crystals from a homogeneous to an inhomogeneous magnetic state is described phenomenologically on the basis of the thermodynamic potential with gradient terms. In the discussion of physical reasons for magnetic order parameter modulation in weak ferromagnetic doped with diamagnetic ions, preference is given to the mechanism associated with the emergence of uniaxial magnetic centers with a random distribution of azimuths of easy axes in the basal plane of the crystal in the vicinity of impurities. A model describing the formation of a modulated magnetic state in α-Fe_2O_3:Ga and $FeBO_3$:Mg is proposed, according to which the competition between magnetoanisotropic and Zeeman interactions in the inhomogeneous magnetic phase of these crystals leads to periodic deviations in the direction of the local ferromagnetism vector from the direction of magnetization. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Intense studies of modulated structures with various order parameters in solid media are associated with the analysis of the mechanisms of formation of such structures and with the search for nontraditional phenomena important for practical applications (e.g., in devices for data recording, processing and storage). This also applies, for example, to dielectrics with dipole and quadrupole order parameters, in which hysteretic optical and acoustic effects attract special attention, and to one-dimensional conductors, in which a pronounced anomaly in the electrical conductivity is observed after the formation of a charge density wave. From this point of view, the magnets with traditionally studied magnetic structures are of considerable interest; for this reason, the main mechanisms leading to modulation of the magnetic order parameter have been studied comprehensively.

It has been established that the main reasons for the instability of a homogeneous magnetic state of magnetically ordered dielectrics are either exchange interactions of various signs between the nearest and next-to-nearest atomic neighbors or the Dzyaloshinskii–Moriya relativistic interaction. The long-period (incommensurate) modulated magnetic structures emerging in this case have been thoroughly studied both theoretically and experimentally (see, for example, [1]). In addition, the presence of a random field may also lead to modulation of the magnetic order parameter of the medium under certain conditions. Such a possibility was obviously indicated was the first time by Imry and Ma [2], who studied theoretically the instability of a homogeneous magnetic state of a ferromagnet induced by an insignificant random field, which makes the state with a nonuniform magnetization advantageous from the energy point of view. The fact that a random magnetic field facilitates the formation of long-period magnetic structure was unexpected. Communication [2] gave rise to a series of publications devoted to random field effects with the model medium in the form of an axial Ising antiferromagnet with random bonds between magnetically active ions. In general, the experimental results confirmed the theoretical concepts and made it possible to determine the critical dimensionality below which the long-range magnetic order is absent in the medium [3], to establish a peculiar effect of the external magnetic field on the critical parameters in the vicinity of phase-transition points [4], and to observe the presence of a micromagnetic structure in the uniaxial doped antiferromagnet MnF_2:Zn [3, 5].

The emergence of a heterogeneous magnetic structure was also predicted for easy-plane antiferromagnets. For example, it was proved theoretically [6] that the conventional domain structure with a uniform antiferromagnetism vector—as well as a structure in which the antiferromagnetism vector continuously changes its orientation, remaining in the basal plane—can be realized in an easy-plane antiferromagnet (depending on the correlation radius of random anisotropic fields) with bulk or point crystal lattice defects. The magnetic structures formed in this class of magnetic dielectrics have not been studied experimentally to our knowledge.

We describe here the results of analysis of the effect of diamagnetic impurity ions on the magnetic state of easy-plane antiferromagnets with weak ferromagnetism as well as the conditions of excitation and existence of modulated magnetic structures in these crystals. The observed magnetic superstructures are described phenomenologically on the basis of the thermodynamic potential with gradient terms; physical mechanisms of the formation of such structures are considered.

2. SAMPLES AND MEASURING TECHNIQUE

Weak ferromagnets occupy a special place among easy-plane antiferromagnets. The existence of the ferromagnetic moment in such materials makes it possible to control their magnetic structure with the help of a weak magnetic field, which makes it possible in principle to easily trace the effects associated with the presence of a random field (associated, for example, with the random nature of exchange or with local fluctuations in the values of the competing anisotropic interactions responsible for the magnetic state of the medium). We chose as objects of investigation the rhombohedral antiferromagnets with weak ferromagnetism (space group D_{3d}^6), viz., hematite α-Fe_2O_3 and iron borate $FeBO_3$ doped with diamagnetic impurities to create crystal lattice defects. We used the magnetooptical method for visualization of the domain structure as the experimental technique for studying the magnetic state of these weak ferromagnets.

The choice of the objects and method of investigation was mainly dictated by the following circumstances: the magnetic, optical, and magnetooptical properties of hematite and iron borate are well known [7] and the behavior of their domain structures in an external magnetic field has been intensely studied [8–10]. This enables us to thoroughly analyze the difference in the magnetic response of nominally pure (undoped) crystals and crystals diluted with a diamagnetic impurity. In addition, weak ferromagnets with a high Néel temperature T_N are promising media as an elemental basis of various rf devices; the stringent technological criteria of homogeneity of macroscopic magnetic parameters of the medium existing in this field necessitate determining the possible reasons for the magnetic inhomogeneity of the medium.

During the synthesis of hematite and iron borate, gallium and manganese oxides were added, respectively, to the furnace charge in an amount constituting about 0.5% by weight of iron oxide. Synthesized $FeBO_3$:Mg crystals were in the form of 3×3 mm^2 plane-parallel plates 50–60 μm in thickness. The principal symmetry axis of the crystal (C_3 axis) coincided with the normal to the developed plane of the plate. The crystal surface had a high optical quality and was not subjected to any additional treatment. In the case of α-Fe_2O_3:Ga single crystals, 200–300-μm thick plates were cut so that the C_3 axis coincided with the normal to the sample plane. The developed surfaces of the plates were polished with diamond pastes to obtain an optical quality. For our experiments, we chose crystals with a shape close to a regular hexagon in the basal plane. The samples were oriented in the basal plane (by determining the direction of the C_2 axis) from the natural faceting of the synthesized crystals. The samples were annealed in air to remove mechanical stresses and obtain a more uniform distribution of impurity over volume. Diamagnetic measurements proved that diamagnetic impurities virtually did not change the Néel temperature of the synthesized samples as compared to undoped α-Fe_2O_3 ($T_N \approx 950$ K) and $FeBO_3$ ($T_N \approx 350$ K), while doping of hematite with gallium ions shifted its Morin temperature to below the boiling point of liquid nitrogen.

Domain structures were observed in α-Fe_2O_3:Ga and $FeBO_3$:Mg through a polarization microscope equipped with an image converter and a TV camera interfaced with a computer. Experiments were made in the range of the maximal optical transparency of hematite and iron borate (at a wavelength of $\lambda \approx 1$ and 0.5 μm, respectively) in transmitted light with a normal incidence on the sample surface in the crossed geometry of the axes in the polarizer–analyzer system.

The external magnetic field used for studying the evolution of domain structures of the samples was produced by two pairs of Helmholtz coils and was applied in the basal plane of the crystals. The magnetization system enabled us to orient the field vector **H** along any direction in the basal plane of the sample for $|\mathbf{H}| = \text{const}$. The samples were placed in an optical cryostat that made it possible to carry out measurements in the temperature range 80–290 K. In all experiments, the sample plane was oriented at right angles to the magnetic field of the Earth.

When the spatial distribution of the magnetic order parameter is studied in an easy-plane weak ferromagnet using the magnetooptical method, it becomes necessary to establish the relation between the intensity I of light transmitted through the polarizer–sample–analyzer system and the orientation of the ferromagnetism vector **m** (or antiferromagnetism vector **l**) in the basal plane of the crystal. This can be done most conveniently from an analysis of the dependence of intensity I on the orientation of the external magnetic field applied in the plane

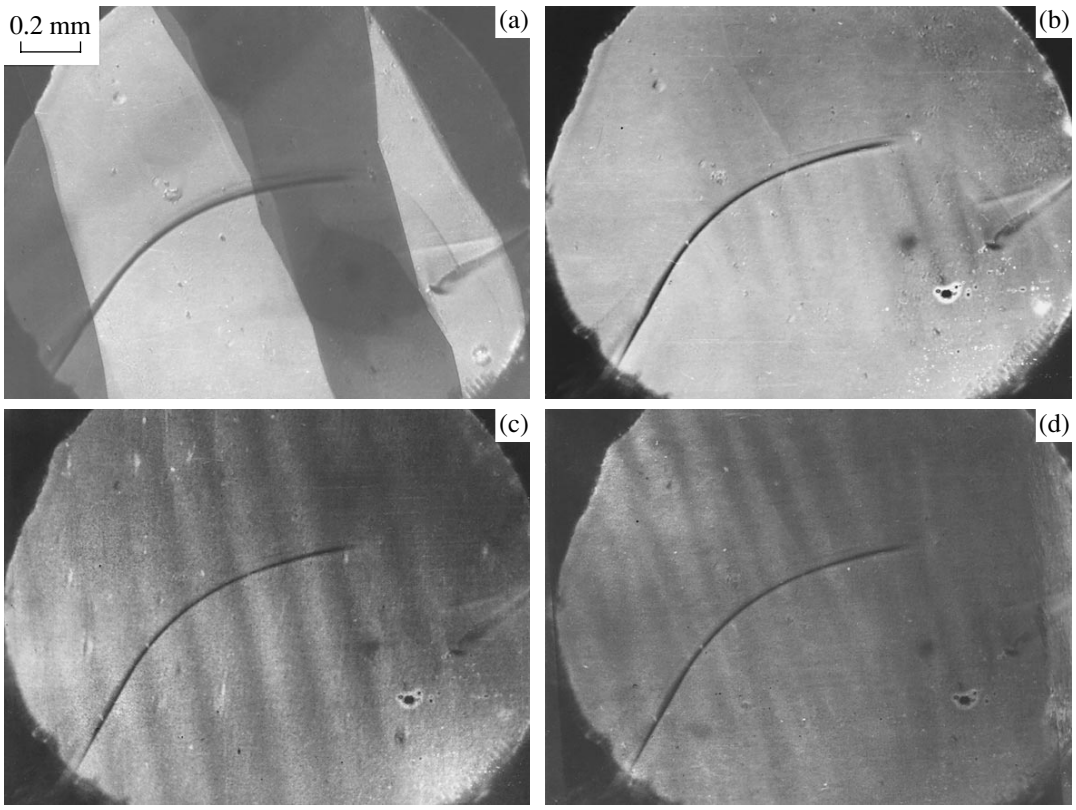


Fig. 1. Images of the $FeBO_3$:Mg sample obtained in polarized light for different values of magnetic field H: (a) 0, (b) 4.5 Oe, (c) 5 Oe, and (d) 10 Oe (T = 80 K, field **H** is perpendicular to the C_2 axis).

of the sample. The magnitude of the applied magnetic field must ensure the homogeneous (monodomain) magnetic state of the samples; i.e., the condition $\mathbf{m} \parallel \mathbf{H}$ must be satisfied. It was found from the measurement of the field dependences of the magnetooptical signal I that the latter condition was satisfied for $H > 30$ Oe.

Considering that the main magnetooptical effect in iron borate and hematite is the linear magnetic dichroism [11] and linear magnetic birefringence [8, 12], respectively, in the spectral range with the maximal optical transparency for light propagating along the optical axis (C_3), in the crossed geometry of the polarizer and analyzer transmission axes, the light intensity at the exit of the polarizer–sample–analyzer system in the case of $FeBO_3$:Mg can be represented in the form

$$I(\varphi) \approx I_0 S \sin 2(\varphi - \Psi); \quad (1)$$

in the case of a α-Fe_2O_3:Ga sample, the corresponding expression has the form [12]

$$I(\varphi) \approx I_0 P[1 - \cos 4(\varphi + \Psi)], \quad (2)$$

where I_0 is the intensity of incident light, Ψ is the azimuth of the transmission axis of the polarizer relative to the x axis of the laboratory system of coordinates (in which the z axis is parallel to C_3), φ is the azimuth of the ferromagnetism vector relative to the same axis in the plane perpendicular to the direction of light propagation (we assume that $\mathbf{m}(Z)$ = const in this case), S is the magnetooptical coefficient characterizing linear magnetic dichroism, and P is the linear magnetic birefringence (Cotton–Mouton effect).

The results of our measurements proved that the light intensity I varies with the orientation of vector **H** ($H > 30$ Oe) in the basal plane of the crystal samples virtually in accordance with formulas (1) and 2 (the $I(\varphi)$ dependence for α-Fe_2O_3:Ga is similar to that in [13]).

Thus, an analysis of the local characteristic of luminous flux at the sample exit based on formulas (1) and (2) gives the idea of spatial variations of the magnetic order parameter in $FeBO_3$:Mg and α-Fe_2O_3:Ga if we assume that angle φ is a function of spatial coordinates.

3. EXPERIMENTAL RESULTS

To study the magnetic state of $FeBO_3$:Mg and α-Fe_2O_3:Ga crystals in the course of magnetization, we visualized the evolution of their domain structures. In the demagnetized state, the samples had a domain structure typical of iron borate and hematite crystals (see, for example, [8–10]) with domain walls orienta-

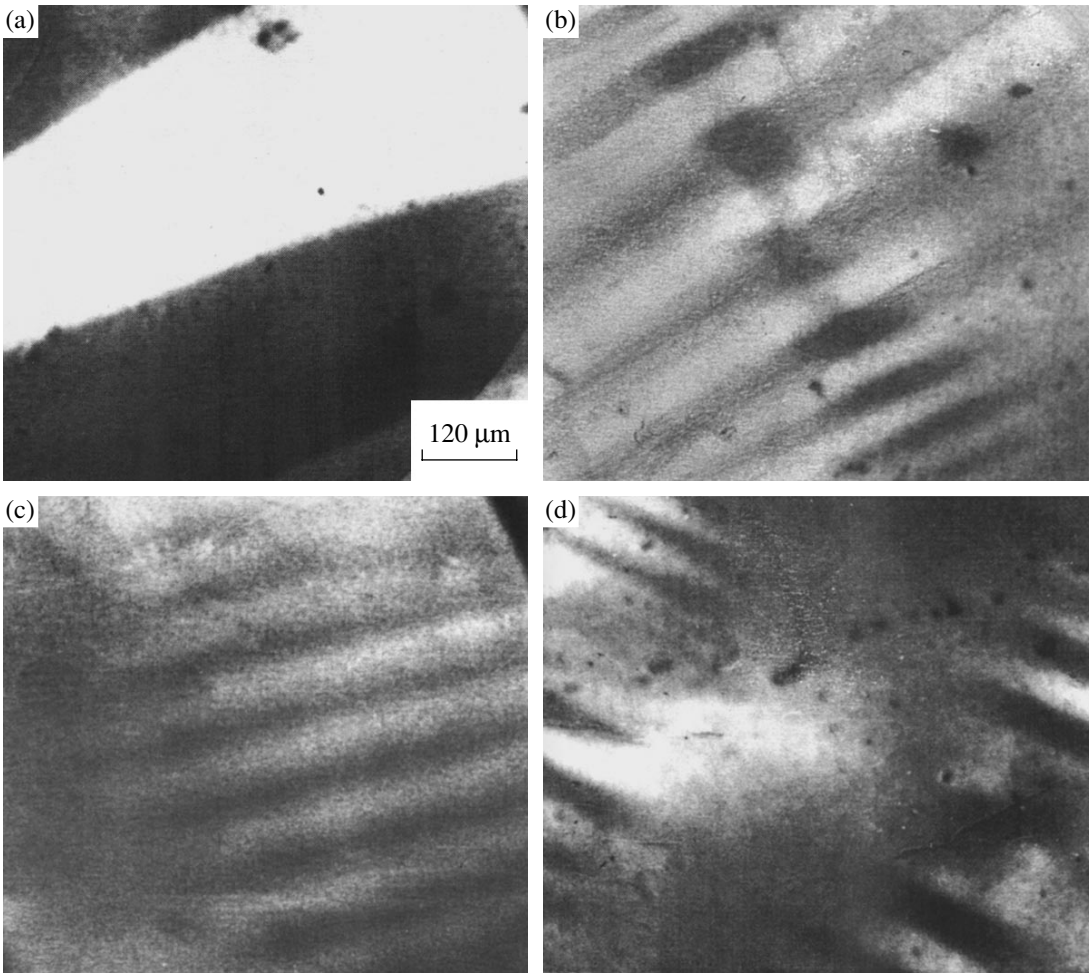


Fig. 2. Images of the α-Fe_2O_3:Ga sample obtained in polarized light at T = 290 K in fields H = 0 (a) and 7 Oe (b–d) (the direction of vector **H** in each frame is perpendicular to one of the C_2 axes).

tion in directions close to C_2 axes (Figs. 1a and 2a show the images of the crystal samples obtained in polarized light for H = 0). The process of technical magnetization of $FeBO_3$:Mg and α-Fe_2O_3:Ga samples was traditional: the area of domains in which vector **m** forms an acute angle with vector **H** increased with the field at the expense of domains with the opposite orientation of **m** until the crystal passed to the monodomain (homogeneous) state. However, in a certain temperature range, the application of a magnetic field in the plane of the sample in directions perpendicular to C_2 axes (i.e., along hard axes of intraplanar crystallographic anisotropy [8–10]) first imparts a monodomain structure to the crystals (Fig. 1b). Upon a further increase in the field, the domain image acquires fringes with blurred boundaries (Figs. 1c, d and 2b–d). It was found that the observed system of fringes with various magnetooptical contrast in α-Fe_2O_3:Ga appeared in the field interval 6 Oe < H < 16 Oe and abruptly disappeared as the temperature is lowered from room temperature to $T_k \approx$ 260 K. In $FeBO_3$:Mg, the system of fringes existed below $T_k \approx$ 135 K in a certain field interval depending on T and visually disappeared when the magnetooptical contrast between bright and dark fringes was gradually reduced.[1] It should be noted that domain structures observed in both crystals for H = 0 in the temperature range where system of fringes exists are virtually independent of T.

Figure 3 shows the experimentally obtained diagram containing the regions of observation of a quasiperiodic system of fringes with different magnetooptical contrasts as a function of the applied field and its direction in the basal plane of the $FeBO_3$:Mg sample (for α-Fe_2O_3:Ga, the directional diagram for the emergence of fringes is analogous, the only difference being the

[1] Analogous experiments performed on nominally pure (undoped) crystals of hematite and iron borate showed that, under the conditions of technical saturation of magnetization (H > 5 Oe), manipulations with the magnitude and direction of the external magnetic field did not lead to the formation of a periodic modulation in the contrast of the images of these crystals obtained in polarized light.

ranges of fields and temperatures in which the diagram is realized [14]). Hatched regions of the diagram correspond to field values and azimuth for which the modulation of the sample image contrast is observed in polarized light. The direction of hatching corresponds to the direction of fringes with various magnetooptical contrasts.

It was found that the mean spatial period d of the system of fringes formed in the sample image depends on the magnitude of the applied field and (for $FeBO_3$:Mg) on temperature. Figures 4 and 5 show the dependences of the period d of the system of fringes observed for $FeBO_3$:Mg on the temperature and the magnetic field applied along the normal to its wavefront. The variation of d as a function of H and T occurs jumpwise due to a change in the number of fringes per unit length (this is shown by steps in Figs. 4 and 5). It should be noted that, in addition of this pinning effect, the field dependence of the period of the system of fringes displays hysteresis in the values of d: with decreasing H, the mean period of the system of fringes changes to a lesser extent than upon an increase in the field from zero (the $d(H)$ dependence obtained for α-Fe_2O_3:Ga in [14] resembles qualitatively the dependence in Fig. 5).

Let us consider physical reasons for the emergence of a system of fringes with various magnetooptical contrast, which are observed in $FeBO_3$:Mg and α-Fe_2O_3:Ga in polarized light. These fringes may be (i) the interference fringes appearing as a result of superposition of the Faraday effect and natural birefringence in Bloch domain walls separating domains over the sample thickness [15, 16]; (ii) the images of domains with Néel-type domain walls inclined to the basal plane [8]; (iii) the image of strip domain structure; and (iv) the fringes may reflect the presence of a spatially modulated (inhomogeneous) magnetic state.

It follows from the results described in [15, 16] that interference fringes on the surface of a $FeBO_3$ crystal are observed only for quite large angles of incidence (about 45°) of light on the sample surface and disappear when light propagates along the C_3 axis (i.e., in the geometry used in our experiments). In addition, no system of fringes similar to those shown in Figs. 1 and 2 emerged under the same conditions of magnetization in nominally pure crystals of hematite and iron borate.

The presence of Néel domain walls inclined to the basal plane of the crystals is visualized as fringes with blurred boundaries [8]. However, when the sample is rotated around the axis lying in the basal plane along the direction of the fringes, the areas of bright and dark fringes must change (due to a change in the projection of domain walls onto the plane of the sample), which is not observed in experiments.

As regards the possibility of existence of a strip structure, we must assume that impurities in α-Fe_2O_3:Ga and $FeBO_3$:Mg considerably alter the hexagonal anisotropy constant, ensuring the emergence of vector **m** from the basal plane (in this case, the sample must be split into strip domains to decrease the magnetostatic energy). For this purpose, the sixth-order anisotropy constant in doped crystals (in contrast to nominally pure crystals) must have an appreciable value as compared to the second-order anisotropy constant, which is hardly probable since the crystals are doped with diamagnetic ions.

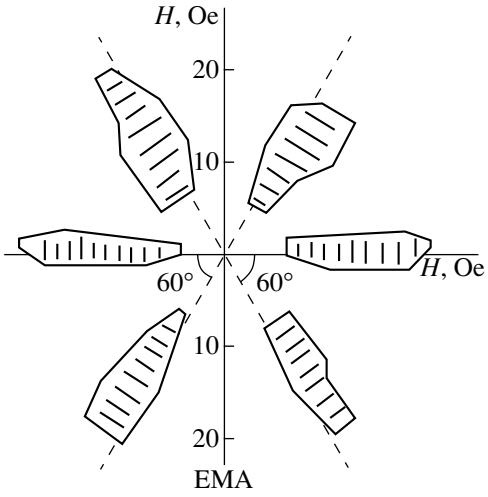


Fig. 3. Diagram for the existence of structure of fringes with different magnetooptical contrasts in $FeBO_3$:Mg on the plane specified by the azimuth of vector **H** and the applied field (T = 80 K). Hatched regions correspond to the values of the field and its azimuth, for which the emergence of the structure is observed; the orientation of hatching corresponds to the orientation of the wavefronts of the structure. The easy magnetization axis (EMA) coincides with the direction of one of the C_2 axes.

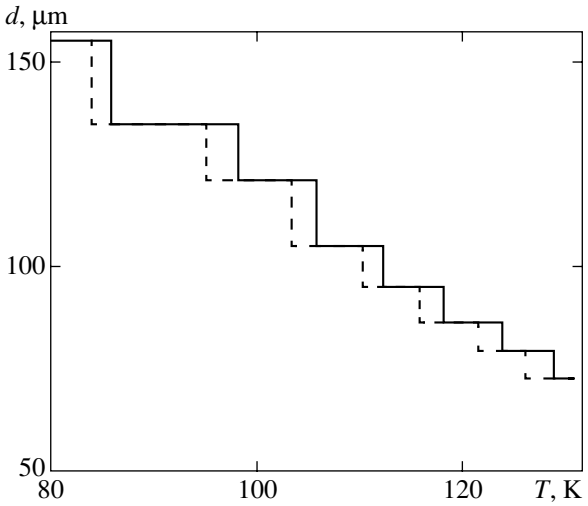


Fig. 4. Temperature dependence of the spatial period of the system of fringes observed in $FeBO_3$:Mg during sample heating (solid line) and cooling (dashed line) (H = 6 Oe, field **H** is perpendicular to the C_2 axis).

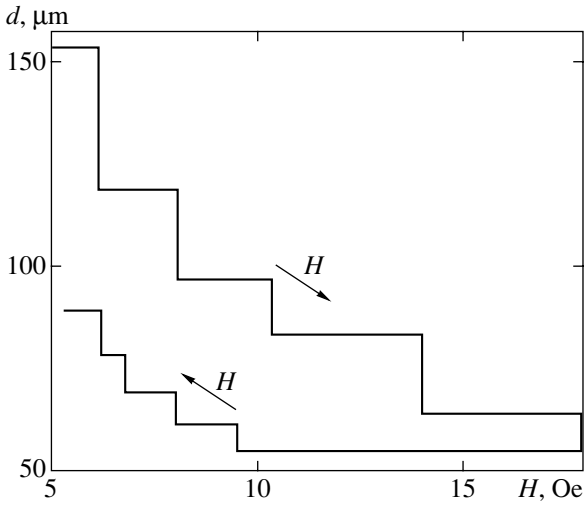


Fig. 5. Field dependence of the spatial period of the system of fringes with different magnetooptical contrasts, observed in $FeBO_3$:Mg at $T = 80$ K.

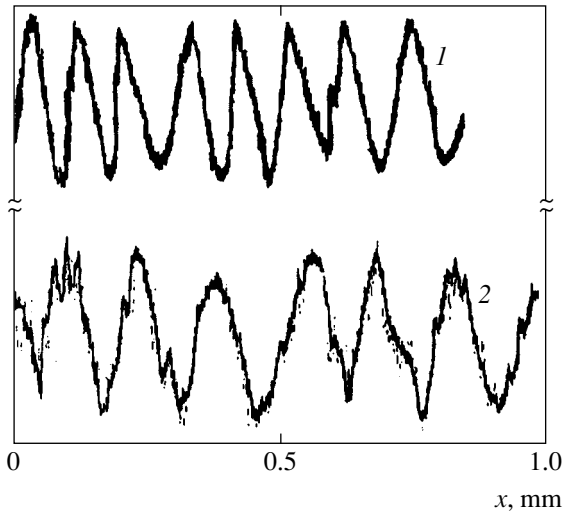


Fig. 6. Dependence of the magnetooptical signal I on the spatial coordinate along the direction of the wave vector of a modulated magnetic structure, observed in α-Fe_2O_3:Ga (*1*) and $FeBO_3$:Mg (*2*).

Thus, it remains for us to assume that weak ferromagnets in fields H exceeding the saturation field exhibit a transition from a state with uniform magnetization to a modulated magnetic state. The absence of sharp boundaries in the image of fringes with various magnetooptical contrast indicates that, unlike the ordinary domain structure for which the azimuth of the magnetization vector is constant within an individual domain, the ferromagnetism (antiferromagnetism) vector in the magnetic states realized in our experiments changes its azimuth continuously and smoothly along the direction of magnetization. Computer digitization of the images of the observed systems of fringes showed that modulation of intensity I along the direction perpendicular to the direction of the fringes is correctly described by a harmonic function of a coordinate (the $I(x)$ dependences obtained in this way with field **H** parallel to the x axis and perpendicular to the C_2 axis) for α-Fe_2O_3:Ga and $FeBO_3$:Mg are shown in Fig. 6). If we assume that vector **m** in modulated magnetic structures does not emerge from the basal plane of the crystal, we can determine local variations of its azimuth in the observed system of fringes using Eqs. (1) and (2). This can be done experimentally by rotating the polarization plane of light incident on the sample relative to the direction of the fringes in the structure and determining the position of the polarizer, for which the images of bright and dark fringes are inverted. According to the results of our measurements, the latter effect was obtained for α-Fe_2O_3:Ga in the field $H = 6$ Oe by rotating the polarizer through an angle of about 30° relative to the position in which the fringes shown in Figs. 2b–d are observed. To within the experimental error, the magnitude of the angle is independent of T in the entire temperature range of modulated magnetic structures. In accordance with expression (2), this means that the amplitude φ_A of modulation of the direction of vector **m** in such structures in α-Fe_2O_3:Ga for $H = 6$ Oe amounts approximately to 15° and is temperature-independent in the region from T_k to $T = 290$ K. Analogously, it was found that the maximal amplitude φ_A of modulation of the azimuth of vector **m** in the structure for $FeBO_3$:Mg at $T = 80$ K amounts to approximately 12° and decreases upon heating (Fig. 7). The field dependences of angle φ_A for both investigated crystals are shown in Fig. 7.[2]

On the basis of the obtained results, the modulated magnetic state of α-Fe_2O_3:Ga and $FeBO_3$:Mg can be visualized in the form of a 1D dependence of azimuthal angle φ characterizing the orientation of vector **m** in the basal plane on the spatial coordinate along the direction of magnetization (field **H** is perpendicular to the C_2 axis). The components of vectors **m** and **l** in modulated magnetic structures in a coordinate system in which the z axis coincides with the direction of light propagation and the x axis coincides with the direction of the applied field can be defined as

$$\begin{aligned} l_x &= |\mathbf{l}|\cos\varphi_A\cos ky, \quad m_x \approx |\mathbf{m}|, \\ l_y &\approx |\mathbf{l}|, \quad m_y = |m|\cos\varphi_A\cos kx, \end{aligned} \tag{3}$$

where k is the wave vector of the structure.

[2] Judging from the obtained dependences $\varphi_A(T)$, the transition from a homogeneous to a modulated magnetic state is a first-order transition in α-Fe_2O_3:Ga and a second-order transition in $FeBO_3$:Mg; the transition in modulated magnetic structures occurs above the boundary temperature T_k in the former case and below this temperature in the latter case.

According to visual perception and the behavior in a magnetic field, the modulated structures studied here resemble the magnetic superstructure formed under the action of light in iron borate doped with nickel ions [17]. It follows from the theory of a photoinduced modulated magnetic structure proposed in [17] that the exchange interaction between the complexes formed by the matrix and impurity ions in the $FeBO_3$:Ni crystal leads to excitation of the structure; in the absence of photoexcitation, this excitation is small and is enhanced under the action of light. This theory permits in principle the emergence of a modulated magnetic structure even without the participation of light; however, illumination must affect the parameters of the modulated magnetic state formed in the crystal. Since the effect of light on the sample always took place in our experiments, we studied the effect of illumination on the magnetic structure of α-Fe_2O_3:Ga and $FeBO_3$:Mg samples. In our experiments, we varied the power and spectral composition of luminous flux incident on the crystals: introducing one-by-one light filters into the optical channel of the microscope, we cut off various parts of the probe light spectrum (the source of radiation was a 100-W incandescent lamp). However, the variation of the luminous flux density and light spectrum affected neither the excitation conditions nor the period and direction of the system of fringes; the motion of modulated magnetic structures of the investigated crystals as a whole was not observed either (in contrast to that observed in $FeBO_3$:Ni). Scanning of the sample surface by a laser beam (λ = 0.63 μm, a power on the order of 10 mW) focused into a spot with a diameter of ~20 μm also did not noticeably affect such structures. This means that the excitation of modulation of the azimuth of vector **m** in α-Fe_2O_3:Ga and $FeBO_3$:Mg occurs without the participation of light; in other words, modulated magnetic structures appear in our case for reasons that are not associated with local variations of the exchange interaction, caused by the presence of impurity ions in the composition of the crystals.

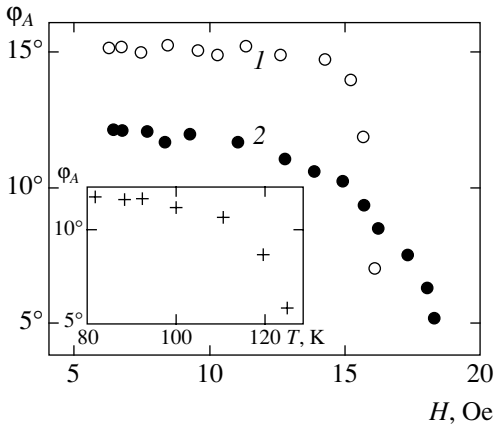


Fig. 7. Field dependence of the amplitude of modulation of the azimuth of the local ferromagnetism vector in a modulated magnetic structure for α-Fe_2O_3:Ga (*1*) at T = 290 K and $FeBO_3$:Mg (*2*) at T = 80 K. The inset shows the temperature dependence of the amplitude of modulation of the azimuth of the local ferromagnetism vector in a modulated magnetic structure observed in $FeBO_3$:Mg (H = 6 Oe).

4. DISCUSSION

To describe the modulated magnetic structure formed in the crystals, we consider the theory of transition of a magnetically ordered medium from a homogeneous to an inhomogeneous magnetic state under the action of a magnetic field [18, 19]. On the basis of experimental facts, we assume that the state in which the local ferromagnetism vector **m** is deflected from the direction of **H** through a certain angle may be stable in an external magnetic field. Since the basal plane of the crystals in question has three directions near which a modulated magnetic structure is formed, we direct the external field **H** for definiteness along one of such directions, which will be treated as the *x* axis. Following [18, 19], we assume that the contributions to the thermodynamic potential density of the crystal from the weakly ferromagnetic moment are insignificant and the potential is determined only by the antiferromagnetism vector components. Then, choosing a small angle β as a formal order parameter characterizing the deviation of the local antiferromagnetism vector **l** from the *y* axis (the *y* axis is parallel to the C_2 axis and the *z* axis is parallel to the C_3 axis), we represent the thermodynamic potential of the crystal, following [18, 19], in the form of the following invariant expansion into a power series in the order parameter:

$$\Phi(\beta) = \int\left[-\frac{1}{2}A\beta^2 + \frac{1}{4}B\beta^4 + \frac{1}{2}\alpha\beta'^2 + \frac{1}{4}\gamma\beta''^2 + mh\beta + \frac{1}{2}\mu lh\beta'^2 + \ldots\right]dy. \quad (4)$$

Here, A, B, α, γ, and μ are parameters independent of H and the primes on β indicate the corresponding derivative with respect to the argument.

The external magnetic field is taken into account in expression (4) by two terms allowed by symmetry. The first term, $mh\beta = mMH\beta$, where M is the sublattice moment, is the Zeeman contribution to the crystal energy; the second term, $\mu lh\beta'^2/2$, is invariant to space and time inversion. Functional Φ constructed in this way differs from that used in [19] only in the presence of this additional term, which takes into account the magnetic inhomogeneity of the medium (the presence of a random field) and renormalizes the coefficient of the square of the first derivative, converting it into a function of H.

A phase transition from the homogeneous magnetic state to a modulated state occurs when the coefficient of the squared derivative in formula (4) becomes smaller than zero. In other words, for $\mu > 0$, in a field $h > \alpha/\mu l$

a modulated magnetic state of the crystal will be induced; in this case, functional (4) is minimized by a function of the form [19]

$$\beta(y) \approx \beta_0 + \eta \cos ky.$$

Thus, in accordance with the model considered here, the application of an external magnetic field in the basal plane of a rhombohedral weak ferromagnet along its hard magnetization axis in a field with the critical value $H_0 = \alpha/M\mu l$ induces a phase transition from the homogeneous magnetic state to a modulated state. The axis along which the modulation emerges in the crystal is oriented along field **H**. The magnetic superstructure can be presented in the form of a ripplon phase in which the azimuth of the local antiferromagnetism (ferromagnetism) vector experiences oscillations with a period $d = 2\pi/k$ against the background of a constant deflection from the given axis. This agrees in principle with the results of measurements (see relations (3)) of the spatial variation of the azimuth of vector **m** in the observed structure if we set $\eta = \varphi_A + \pi/2$ (deviation constant β_0 was not observed experimentally probably because of the error in measuring angle φ_A, which amounted to ~1°).

We can prove that in fields close to the critical value, the characteristics of the emerging inhomogeneous magnetic state behaves as follows:

$$d = 2\pi\sqrt{\frac{2\gamma}{|\alpha + \mu lh|}},$$

$$\beta_0 = \frac{4\gamma h}{M(\alpha + \mu lh)^2},$$

$$\eta^2 = \frac{1}{3B}\left[A + \frac{\alpha + \mu lh}{4\gamma} - \frac{48\gamma^2 h^2 B}{M^2(\alpha + \mu lh)^4}\right].$$

Let us consider the physical meaning of the relations derived on the basis of potential (4). Since the formation of a modulated magnetic state was not observed in nominally pure hematite and iron borate under the same conditions of magnetization, this indicates a direct influence of the impurity on the process of rotation of the sublattice moments in crystals doped with diamagnetic ions during their magnetization in the basal plane. This suggests the presence of anisotropic magnetic centers in α-Fe_2O_3:Ga and $FeBO_3$:Mg in the vicinity of impurity ions; the magnetic moments M_c of these centers are deflected from the directions specified by the intraplanar hexagonal anisotropy of the matrix. In particular, the role of such centers can be played by complexes containing a diamagnetic impurity ion and iron ions located in the coordination sphere closest to it. We can expect that the substitution of diamagnetic impurities for matrix iron ions does not lead to crystallographic symmetry of such magnetic centers that differ from the hexagonal symmetry; however, local distortions of the crystal lattice associated with the difference in the sizes of the main matrix iron ion and magnesium and gallium ions cause considerable changes in the intraplanar anisotropy constant in a certain region in the crystal near impurities. As a result of the presence of this random anisotropy, the orientation of vector M_c of an individual center in the general case does not coincide with the direction of the resultant spontaneous magnetic moment of the entire crystal. Taking into account the fact that the impurity concentration is small and, hence, the magnetic moments of the centers are not ordered cooperatively, we can assume that vectors M_c are oriented along the easy axes of the centers (along three crystallographically equivalent directions in accordance with the chosen model of a center), which are fixed relative to the symmetry axes of the matrix, with statistically equal probabilities. Obviously, after the application of an external magnetic field, the equilibrium magnetic structure of the crystal is determined by the result of the competition between the contributions from random anisotropy, crystallographic anisotropy, and field H to the thermodynamic potential. A similar situation was analyzed in [20] for polycrystalline permalloy films, where the role of random anisotropy was played by crystallographic anisotropy in crystallites oriented at random relative to one another, while the orienting factors were the induced anisotropy (texture) and an external magnetic field. Using the results obtained in [20], we can write the expression for the modulation period in the form

$$d = 2\pi\sqrt{\frac{2J}{M_s H - 2K}}, \quad (5)$$

where J is the exchange constant, M_s is the spontaneous magnetization, and K is the intraplanar hexagonal anisotropy constant (in [20], the constant of induced uniaxial anisotropy was taken for K). It follows hence that a modulated magnetic state can exist in the field region

$$H > H_0 = 2K/M_s. \quad (6)$$

It is important to note that relations (5) and (6) coincide with the expressions for d and H_0 derived above from analysis of potential (4) if we set $J = \gamma$, $2K = -\alpha$, and $M_s = -\mu lM$.

The structure of formula (5) shows that first, for $H \geq H_0$, dependence $d(H)$ decreases sharply; then, for $H > 2H_0$, the modulation period changes insignificantly, which correlates with the experimental data presented in Fig. 5 (for H_0, we assume the minimal field for which a modulated magnetic state is formed). It should be

noted that the periods d calculated by formula (5) using the experimentally obtained values of H_0, as well as the known values of M_s and constant J for both crystals, agree in order of magnitude with the experimentally measured periods (the results of calculation for hematite are given in [13]).

It follows from relation (5) that the temperature dependence of the modulation period is determined by the temperature variation of constant K.[3] We are not aware of the form of dependences $K(T)$ in the crystals studied here. However, if we assume that these dependences are close to the corresponding dependences of the hexagonal anisotropy constant in nominally pure iron borate (where the value of K decreases with increasing T [21]) and hematite ($K \approx$ const for 260 K < T < 300 K [22]), the period of the structure, in accordance with relation (5), should remain unchanged in the entire temperature range of its existence in α-Fe_2O_3:Ga, while the value of d in $FeBO_3$:Mg should decrease as $T \longrightarrow T_k$, which corresponds to the experimental results.

The assumption concerning the presence of uniaxial magnetic centers in the crystals studied here makes it possible to explain the stepwise variation of the spatial period of a modulated magnetic structure as a function of magnetic field and temperature (this concerns only $FeBO_3$:Mg): upon a variation of H and/or T, the number of centers for which the direction of vector $\mathbf{M}_c$ differs from the direction of $\mathbf{H}$ changes discretely in view of discreteness of the azimuths of the easy axes of the centers. This in turn leads to stepwise variation of the ratio between the energies of the competing interactions responsible for the equilibrium magnetic state of the crystal. The difference in the behavior of the dependences $d(H)$ and $d(T)$ observed in $FeBO_3$:Mg (see Section 3) is determined by the difference in the corresponding dependences of concentrations of centers with $\mathbf{M}_c \parallel \mathbf{H}$. At T = const, the difference in the number of centers with $\mathbf{M}_c \parallel \mathbf{H}$ for the direct (upon an increase in the field from zero) and reverse course of magnetization causes a noticeable hysteresis in the dependence $d(H)$ (see Fig. 5). At the same time, the variation of d as a function of T is determined by the temperature dependence of the anisotropy constant specifying the direction of vector $\mathbf{M}_c$ of a center. As a consequence, the values of d measured during heating and cooling of the sample do not differ substantially (the discrepancies in the values of temperature corresponding to jumps in the value of d observed in Fig. 4 are apparently due to error in measuring T, which emerges when the regime of heating is replaced by cooling due to the inertia of the process of stabilization of thermodynamic equilibrium in the crystal).

Thus, the proposed model for the emergence of an inhomogeneous magnetic state makes it possible to explain, at least qualitatively, the experimentally observed behavior of the period of a modulated magnetic structure in α-Fe_2O_3:Ga and $FeBO_3$:Mg depending on temperature and the applied magnetic field (except at boundary values of temperature T_k). According to Hoffmann [20], an inhomogeneous magnetic state of the medium emerges for $K_a > K \gg K_a N^{-1/2}$, where N is the number of anisotropic magnetic centers per unit volume and K_a is the anisotropy constant defining the direction of the easy axis of the center. This suggests that the temperature range where modulated magnetic structures exist in the crystals considered here is determined by the relation between the competing magnetoanisotropic interactions, which depends on T.

5. CONCLUSIONS

It has been found that magnetic inhomogeneities associated with quasiperiodic deviations of the ferromagnetism vector from the direction of magnetization appear in easy-plane weak ferromagnets doped with diamagnetic ions in a magnetic field exceeding the saturation field for nominally pure crystals. The mechanism of formation of a modulated magnetic state in this class of magnets presumes the emergence of anisotropic magnetic centers with a random distribution of azimuths of easy axes in the basal plane of the crystal in the vicinity of diamagnetic impurities. The direction of the induced magnetic superstructure, as well as the spatial period, varies depending on the orientation and magnitude of the external magnetic field; this may be of interest for possible practical applications of this effect.

[3] According to the results of measurements, in the temperature range of the modulated magnetic state, I/I_0 = const for both crystals under investigation to within the experimental error; consequently, dependence $M_s(T)$ in relation (5) can be disregarded.

REFERENCES

1. Yu. A. Izyumov, Usp. Fiz. Nauk **144**, 439 (1984) [Sov. Phys. Usp. **27**, 845 (1984)].
2. Y. Imry and S. Ma, Phys. Rev. Lett. **35**, 1399 (1975).
3. R. J. Birgeneau, H. Yoshizawa, R. A. Cowley, *et al.*, Phys. Rev. B **28**, 1438 (1983).
4. H. Ikeda, J. Phys. C **16**, L21 (1983).
5. H. Ikeda, J. Phys. C **16**, L1033 (1983).
6. E. B. Sonin, J. Phys. C **13**, 3293 (1980).
7. G. A. Smolenskiĭ, V. V. Lemanov, G. M. Nedlin, *et al.*, *The Physics of Magnetic Dielectrics* (Nauka, Leningrad, 1974) [in Russian].
8. V. L. Preobrazhenskiĭ, A. A. Shishkov, and N. A. Ékonomov, Fiz. Tverd. Tela (Leningrad) **29**, 3549 (1987) [Sov. Phys. Solid State **29**, 2034 (1987)].
9. G. B. Scott, J. Phys. D **7**, 1574 (1974).
10. J. A. Eaton and A. H. Morrish, Can. J. Phys. **49**, 2768 (1971).

11. Yu. M. Fedorov, A. A. Leksikov, and A. E. Aksenov, Fiz. Tverd. Tela (Leningrad) **26**, 220 (1984) [Sov. Phys. Solid State **26**, 128 (1984)].

12. I. A. Merkulov, E. G. Rudashevskiĭ, H. LeGall, and C. Leicuras, Zh. Éksp. Teor. Fiz. **75**, 628 (1978) [Sov. Phys. JETP **48**, 316 (1978)].

13. A. T. Karaev and B. Yu. Sokolov, Zh. Tekh. Fiz. **73** (5), 130 (2003) [Tech. Phys. **48**, 651 (2003)].

14. Z. T. Azamatov, A. T. Karaev, B. Yu. Sokolov, and Yu. M. Fedorov, Zh. Tekh. Fiz. **71** (3), 84 (2001) [Tech. Phys. **46**, 354 (2001)].

15. J. Haisma and W. T. Stacy, J. Appl. Phys. **44**, 3367 (1973).

16. A. V. Chzhan, Fiz. Tverd. Tela (Leningrad) **32**, 280 (1990) [Sov. Phys. Solid State **32**, 158 (1990)].

17. Yu. M. Fedorov, A. F. Sadreev, and A. A. Leksikov, Zh. Éksp. Teor. Fiz. **93**, 2247 (1987) [Sov. Phys. JETP **66**, 1283 (1987)].

18. A. Michelson, Phys. Rev. B **16**, 585 (1977).

19. I. E. Dikshteĭn, F. V. Lisovskiĭ, E. G. Mansvetova, and V. V. Tarasenko, Fiz. Tverd. Tela (Leningrad) **25**, 2545 (1983) [Sov. Phys. Solid State **25**, 1465 (1983)].

20. H. Hoffmann, J. Appl. Phys. **35**, 1790 (1964).

21. V. F. Doroshev, I. M. Krygin, and S. N. Lukin, Pis'ma Zh. Éksp. Teor. Fiz. **29**, 286 (1979) [JETP Lett. **29**, 257 (1979)].

22. G. A. Petrakovskiĭ, A. I. Pankrats, V. M. Sosnin, and V. I. Vasil'ev, Zh. Éksp. Teor. Fiz. **85**, 691 (1983) [Sov. Phys. JETP **58**, 403 (1983)].

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 99, No. 2, 2004, pp. 427–442.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 2, 2004, pp. 483–501.

NONLINEAR PHYSICS

Dynamics of NLS Solitons Described by the Cubic–Quintic Ginzburg–Landau Equation

M. N. Zhuravlev and N. V. Ostrovskaya*

Moscow State Institute of Electronic Engineering (Technical University), Zelenograd, Moscow, 103498 Russia

**e-mail: ost@miee.ru*

Received October 28, 2003

Abstract—The generalized moment method is applied to average the Ginzburg–Landau equation with quintic nonlinearity in the neighborhood of a soliton solution to the nonlinear Schrödinger equation. A qualitative analysis of the resulting dynamical system is presented. New soliton solutions bifurcating from a known exact soliton solution are obtained. The results of the qualitative analysis are compared with those obtained by direct numerical solution of the Ginzburg–Landau equation. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Equations similar to the Ginzburg–Landau equation

$$i\psi_t + \sigma\psi_{xx} + N[|\psi|]\psi = 0, \qquad (1)$$

where t and x are independent variables, $\psi(x, t)$ is a complex-valued function, σ is a real parameter, and $N[|\psi|]$ is a polynomial operator of the form $c_1 + c_2|\psi|^2 + c_3|\psi|^4$ (c_1, c_2, and c_3 are complex parameters), arise in analyses of various phenomena in nonlinear optics, plasma physics, superconductivity theory, and fluid mechanics. When $\sigma = 1/2$ and $N = |\psi|^2$, the Ginzburg–Landau equation (GLE) reduces to the nonlinear Schrödinger (NLS) equation with cubic nonlinearity. The latter is a completely integrable equation having soliton particular solutions.

To date, several families of analytical solutions to GLE (1) have been found (e.g., see [1, 2]). They include families of localized solutions with fixed and arbitrary amplitudes, also called dissipative solitons. Furthermore, some soliton-like numerical solutions to GLE have been reported that fall outside the aforementioned classes and have very special properties (e.g., see [3]). These solutions are special in that they have certain periodic characteristics, such as width, amplitude, or center-of-mass location. One would naturally try to standardize the procedure of finding them by performing a qualitative analysis of some dynamical system obtained by averaging the original equation. All averaging methods are both advantageous and disadvantageous in that they reduce the dynamics determined by the initial distribution to the dynamics of a limited number of parameters of the initial condition. In this paper, we make use of the generalized moment method [4], which is undoubtedly advantageous in that it can be used to perform a separate analysis of the dynamics of both amplitude and width of a soliton solution to the unperturbed problem.

2. BASIC EQUATIONS

The GLE with quintic nonlinearity used in nonlinear optics has the form

$$\begin{aligned} i\psi_t + \sigma\psi_{xx} + \eta|\psi|^2\psi &= i\delta\psi + i\epsilon|\psi|^2\psi \\ &+ i\beta\psi_{xx} + i\mu|\psi|^4\psi - \nu|\psi|^4\psi, \end{aligned} \qquad (2)$$

where δ, β, ϵ, μ, and ν are real constants (which may not be small) and ψ is a complex field. In the special case of an optical fiber, these quantities are interpreted as follows: ψ is the complex envelope of an electromagnetic field; the parameters δ, ϵ, and μ characterize pumping or energy dissipation (depending on their signs); β is a filtering coefficient; and ν is the coefficient of a quadratic correction to a nonlinear refractive index. The parameter σ characterizes second-order dispersion ($\sigma = +1/2$ and $\sigma = -1/2$ correspond to anomalous and normal dispersion, respectively), and $\eta = 1$.

We rewrite the original equation as

$$i\psi_t + \sigma\psi_{xx} + \eta|\psi|^2\psi = R[\psi], \qquad (3)$$

where

$$R = i\delta\psi + i\epsilon|\psi|^2\psi + i\beta\psi_{xx} + i\mu|\psi|^4\psi - \nu|\psi|^4\psi.$$

If the right-hand side is zero, then we have the conventional nonlinear Schrödinger equation with cubic nonlinearity. The first two integrals of the NLS equation (energy and momentum) are

$$I_1 = \int_{-\infty}^{\infty} |\psi|^2 dx, \quad I_2 = \frac{1}{2}\int_{-\infty}^{\infty} (\psi\psi_x^* - \psi^*\psi_x) dx. \qquad (4)$$

Following [4], we define the first two moments of ψ:

$$D_1 = \int_{-\infty}^{\infty} x|\psi|^2 dx, \quad D_2 = \int_{-\infty}^{\infty} (x - x_c)^2|\psi|^2 dx, \qquad (5)$$

where x_c is the center-of-mass coordinate of ψ:

$$x_c = D_1/I_1.$$

The evolution of these four characteristics as functions of t is described by the equations

$$\begin{aligned}
\frac{dI_1}{dt} &= i\int_{-\infty}^{\infty}(\psi R^* - \psi^* R)dx,\\
\frac{dI_2}{dt} &= -i\int_{-\infty}^{\infty}(\psi_x R^* + \psi_x^* R)dx,\\
\frac{dD_1}{dt} &= 2i\sigma I_2 - i\int_{-\infty}^{\infty}x(\psi R^* - \psi^* R)dx,\\
\frac{dD_2}{dt} &= -2i\sigma M_1 - i\int_{-\infty}^{\infty}(x-x_c)^2(\psi R^* - \psi^* R)dx,
\end{aligned} \tag{6}$$

where

$$M_1 = \int_{-\infty}^{\infty}(x-x_c)(\psi^*\psi_x - \psi\psi_x^*)dx.$$

System (6) is supplemented with the identity

$$\begin{aligned}
\int_{-\infty}^{\infty}(\psi\psi_t^* - \psi_t\psi^*)dx &= 2i\int_{-\infty}^{\infty}(\sigma|\psi_x|^2 - \mu|\psi|^4)dx\\
&+ i\int_{-\infty}^{\infty}(\psi R^* + \psi^* R)dx.
\end{aligned} \tag{7}$$

Equations (6) and (7) can be used to derive the equations of evolution for an arbitrary set of five parameters of the exact solution to the NLS equation.

As a trial solution to the NLS equation, we take the soliton solution

$$\psi(x,t) = \frac{A}{\cosh[B^{-1}(x-x_c)]}\exp i[\varphi(t) + C(x-x_c)], \tag{8}$$

where A is the soliton amplitude, B is the soliton width, x_c is the center-of-mass coordinate, and C is the wavenumber. When $R \equiv 0$, the parameter A in (8) has an arbitrary value, and the other two are related to it as follows:

$$A = B^{-1}, \quad \varphi(t) = A^2 t/2.$$

For the stationary NLS soliton with center of mass lying on the axis passing through the origin, it holds that $C = 0$ and $x_c = 0$.

For distribution (8),

$$\begin{aligned}
I_1 &= 2A^2B, \quad I_2 = -2iA^2BC,\\
D_1 &= 2x_cA^2B, \quad D_2 = \frac{1}{6}\pi^2A^2B^3,\\
M_1 &= 0.
\end{aligned} \tag{9}$$

If the parameters are continuous functions of t in the neighborhood of the solution, then it follows from (9) that

$$\begin{aligned}
C' &= I_2'\frac{i}{2A^2B} - I_1'\frac{C}{2A^2B},\\
x_c' &= D_1'\frac{1}{2A^2B} - I_1'\frac{x_c}{2A^2B},\\
A' &= I_1'\frac{3}{8AB} - D_2'\frac{3}{2\pi^2AB^3},\\
B' &= D_2'\frac{3}{\pi^2A^2B^2} - I_1'\frac{1}{4A^2}.
\end{aligned} \tag{10}$$

We find I_1', I_2', D_1', and D_2' by direct calculation of the integrals in (6). For $\psi(x, t)$ given by (8), the results are

$$\begin{aligned}
I_1' &= 4B\left(\delta A^2 + \frac{2}{3}\epsilon A^4 + \frac{8}{15}\mu A^6\right) - 4\beta A^2\left(\frac{1}{3B} + BC^2\right),\\
I_2' &= -4iA^2B\left[\left(C\delta - \frac{1}{B^2}\beta C - \beta C^3\right)\right.\\
&\left. + \frac{2}{3}\epsilon CA^2 + \frac{8}{15}\mu CA^4 - \pi^2\beta CB^2\right],\\
D_1' &= A^2\left(4\sigma BC + 4\delta x_cB - \frac{4\beta x_c}{3B} - 4\beta x_cBC^2\right)\\
&+ \frac{8}{3}\epsilon x_cA^4B + \frac{32}{15}\mu x_cA^6B,\\
D_2' &= \frac{1}{3}A^2B^3\left[\delta\pi^2 + 2\epsilon\left(\frac{1}{3}\pi^2 - 2\right)A^2\right.\\
&\left. + 2\mu A^4\left(\frac{4}{15}\pi^2 - 2\right) - \beta\frac{1}{B^2}\left(\frac{1}{3}\pi^2 - 8\right) - \beta\pi^2C^2\right].
\end{aligned}$$

Substituting these expressions into (10), we obtain the following equations for the parameters A, B, C, and x_c of function (8):

$$\begin{aligned}
A' &= -A\beta\left(\frac{1}{3B^2} + \frac{4}{\pi^2B^2} + C^2\right) + A\delta\\
&+ 2\epsilon A^3\left(\frac{1}{3} + \frac{1}{\pi^2}\right) + 2\mu A^5\left(\frac{4}{15} + \frac{1}{\pi^2}\right),
\end{aligned} \tag{11}$$

$$B' = \frac{4}{\pi^2}B\left(2\beta\frac{1}{B^2} - \epsilon A^2 - \mu A^4\right), \quad (12)$$

$$C' = -\frac{4}{3}\frac{1}{B^2}\beta C, \quad (13)$$

$$x_c' = 2\sigma C. \quad (14)$$

The equation for the phase function φ is used to obtain a closed system for the five parameters of the perturbed equation. It is derived separately from identity (7):

$$\varphi' = \sigma C^2 - \sigma\frac{1}{3B^2} + \frac{2}{3}\eta A^2 + \frac{8}{15}\nu A^4 - \beta\frac{C}{B}. \quad (15)$$

Equations (11)–(14) do not contain φ. Therefore, the system can be analyzed without taking into account Eq. (15), and then this equation can be solved separately. Note also that the parameter ν is not contained in the first four equations. For this reason, the numerical analysis of the original equation presented below was performed for a zero value of this parameter.

3. ANALYSIS OF THE DYNAMICAL SYSTEM

The equilibria of system (11)–(14) are obtained by setting the right-hand sides of these equations to zero simultaneously. It follows directly from Eqs. (13) and (14) that the equilibria exist only if $C = 0$ and $x_c = \text{const}$. This implies that only symmetry-preserving stationary soliton-like solutions to GLE can exist in the neighborhood of the solution to the NLS equation for trial function (8). Thus, it is sufficient to analyze only the equations for the soliton amplitude A and width B:

$$A' = -A\frac{1}{B^2}\beta\left(\frac{1}{3} + \frac{4}{\pi^2}\right) + A\delta + 2\epsilon A^3\left(\frac{1}{3} + \frac{1}{\pi^2}\right) + 2\mu A^5\left(\frac{4}{15} + \frac{1}{\pi^2}\right), \quad (16)$$

$$B' = \frac{4}{\pi^2}B\left(2\beta\frac{1}{B^2} - \epsilon A^2 - \mu A^4\right). \quad (17)$$

To simplify the polynomials on the right-hand sides of these equations, we multiply Eq. (16) by $2A$ and Eq. (17) by $2B$. Introducing the new variables $u = A^2$ and $v = B^2$, we obtain

$$u' = -2\beta a_0\frac{u}{v} + 2u(\delta + 2\epsilon a_1 u + \mu a_2 u^2), \quad (18)$$

$$v' = 4\beta b - 2buv(\epsilon + \mu u), \quad (19)$$

where

$$a_0 = \frac{1}{3} + \frac{4}{\pi^2}, \quad a_1 = \frac{1}{3} + \frac{1}{\pi^2},$$

$$a_2 = 2\left(\frac{4}{15} + \frac{1}{\pi^2}\right), \quad b = \frac{4}{\pi^2}.$$

3.1. Unfiltered Problem ($\beta = 0$)

When $\beta = 0$, the dynamical system reduces to

$$u' = 2u(\delta + 2\epsilon a_1 u + \mu a_2 u^2), \quad (20)$$

$$v' = -2buv(\epsilon + \mu u). \quad (21)$$

The system is easily solved. It has the particular solution

$$u = 0, \quad v = c, \quad \varphi = -\frac{\sigma}{3c}t + \varphi_0$$

and the integral

$$-\frac{1}{b}\ln v + c = \frac{1}{2a_2}\ln|P_1| + \frac{\epsilon(a_2 - a_1)}{\mu a_2^2}\int\frac{du}{(u - u_1)^2 - \Delta_1}, \quad (22)$$

where

$$P_1 = \mu a_2 u^2 + 2\epsilon a_1 u + \delta, \quad u_1 = -\frac{\epsilon a_1}{\mu a_2},$$

$$\Delta_1 = \left(\frac{\epsilon a_1}{\mu a_2}\right)^2 - \frac{\delta}{\mu a_2},$$

and c is an integration constant. The final solution depends on the specific values of the parameters:

$$-\frac{1}{b}\ln v + c = \frac{1}{2a_2}\ln|P_1| + \frac{\epsilon(a_2 - a_1)}{\mu a_2^2|\Delta_1|^{1/2}}\arctan\frac{(u - u_1)}{|\Delta_1|^{1/2}}, \quad \Delta_1 < 0, \quad (23)$$

$$-\frac{1}{b}\ln v + c = \frac{1}{2a_2}\ln|P_1| - \frac{\epsilon(a_2 - a_1)}{\mu a_2^2}\frac{1}{u - u_1}, \quad \Delta_1 = 0, \quad (24)$$

$$-\frac{1}{b}\ln v + c = \frac{1}{2a_2}\ln|P_1| + \frac{1}{2\Delta_1^{1/2}}\frac{\epsilon(a_2 - a_1)}{\mu a_2^2} \times \ln\left|\frac{u - u_1 - \Delta_1^{1/2}}{u - u_1 + \Delta_1^{1/2}}\right|, \quad \Delta_1 > 0. \quad (25)$$

The last equation splits into two:

$$-\frac{1}{b}\ln v + c = \frac{1}{2a_2}\ln|P_1| + \frac{\epsilon(a_2 - a_1)}{\mu a_2^2 \Delta_1^{1/2}}\operatorname{artanh}\frac{u - u_1}{\Delta_1^{1/2}}, \quad (26)$$

$$(u - u_1)^2 < \Delta_1,$$

$$-\frac{1}{b}\ln v + c = \frac{1}{2a_2}\ln|P_1| + \frac{\epsilon(a_2 - a_1)}{\mu a_2^2 \Delta_1^{1/2}}\operatorname{arcoth}\frac{u - u_1}{\Delta_1^{1/2}}, \quad (27)$$

$$(u - u_1)^2 > \Delta_1.$$

The solutions $u_\pm$ are the separatrices between these families.

When $\mu = 0$, the solution is

$$-\frac{1}{b}\ln v + c = \frac{1}{2a_2}\ln|P_1| \quad (28)$$

or

$$v = c(2\epsilon a_1 u + \delta)^{-b/2a_2}. \quad (28a)$$

It defines a one-parameter family of power functions with a negative exponent. The family has two asymptotes:

$$u = -\delta/2\epsilon a_1, \quad v = 0.$$

Now, let us discuss some properties of the phase portrait of system (20), (21). First of all, it occupies the domain $u \geq 0$, $v > 0$ in the uv plane. Hereinafter, this domain is referred to as the domain of the phase portrait. The phase portrait always contains the singular line $u = 0$, which represents the "background" solution to Eq. (2) (zero-amplitude soliton of arbitrary width). The points of the singular line are called *trivial singular points*. Note also that the initial conditions of the problem must lie on the curve $uv = 1$ (NLS soliton curve), because the equations were derived by using the fact that the trial function is an exact solution to the NLS equation. This means that only those orbits are admissible that have common points with the NLS soliton curve.

3.1.1. The case of $\mu = 0$

When $\mu = 0$, there is a saddle point $u_0 = -\delta/2\epsilon a_1$ on the axis $v = 0$ (see solution (28a)). If $\delta = 0$ and $\epsilon \neq 0$, then it lies at the origin. If $\delta \neq 0$ and $\epsilon = 0$, then there is no singular point. Saddle points of this kind correspond to finite-amplitude zero-width solitons. We call them degenerate. Even though they do not belong to the range of the functions u and v and have no physical meaning (represent collapsing solitons), they are important for understanding the topology of the phase portrait and interpreting the results obtained by direct numerical computation of Eq. (2). The corresponding phase portraits are shown in Fig. 1. When the parameters ϵ and μ have opposite signs, the saddle point is in the upper half-plane and the separatrix lies in the domain of the phase portrait (the first quadrant of the phase plane). When $\delta < 0$ and $\epsilon > 0$, the separatrix is incident to the singular point; when $\delta > 0$ and $\epsilon < 0$, it originates from the singular point. In the former case, the direction of a trajectory above the separatrix corresponds to growth of pulse amplitude and self-compression (antidissipative compressive dynamics). Below the separatrix, the trajectories terminate on the singular line $u = 0$ (dissipative compressive pulse dynamics against a stable background) (see Fig. 1a). When $\delta > 0$ and $\epsilon < 0$, the phase-portrait topology remains invariant, but the trajectories reverse direction, which reflects the invariance of system (20), (21) under inversion:

$$\delta \longrightarrow -\delta, \quad \epsilon \longrightarrow -\epsilon,$$

$$\mu \longrightarrow -\mu, \quad t \longrightarrow -t,$$

Furthermore, the trajectories originating from the singular line $u = 0$ asymptotically approach the separatrix $u = u_0$ from below (the background is unstable), while the trajectories lying above the separatrix asymptotically approach it from above. In both cases, the variable v increases indefinitely (antidissipative dispersive and dissipative dispersive dynamics, respectively) (see Fig. 1b). In this system, dissipative and pumping processes cannot balance out if $\mu = 0$.

When δ and ϵ have similar signs, the saddle point is located in the lower half-plane, i.e., outside the domain of the phase portrait. When δ and ϵ are positive, all trajectories lying in the first quadrant approach a vertical separatrix (antidissipative compressive pulse dynamics). When δ and ϵ are negative, all trajectories terminate on the singular line (dissipative dynamics) (see Fig. 1c). In the former case, energy is only pumped into the system and the initial disturbance grows indefinitely. In the latter case, energy is only dissipated and the pulse disappears completely.

When $\delta = 0$, the saddle point is at the origin and the singular line separates two families of similarly directed hyperbolic trajectories. When $\epsilon < 0$, the system exhibits dissipative dispersive dynamics (see phase portrait in Fig. 1d). When $\epsilon > 0$, antidissipative compressive dynamics are observed.

When $\delta \neq 0$ and $\epsilon = 0$, the trajectories are vertical lines that are either incident on or originate from the singular line $u = 0$, depending on the sign of δ (see Fig. 1e).

When both $\delta = 0$ and $\epsilon = 0$, the domain of the phase portrait is a plane consisting of singular points, except for the points not belonging to the NLS soliton curve. Thus, the phase portrait reduces to the singular curve $uv = 1$ and the isolated singular point (0, 0) (see Fig. 1f).

3.1.2. The case of $\mu \neq 0$

In this case, the system has equilibria other than trivial or degenerate ones. They are found by setting to zero

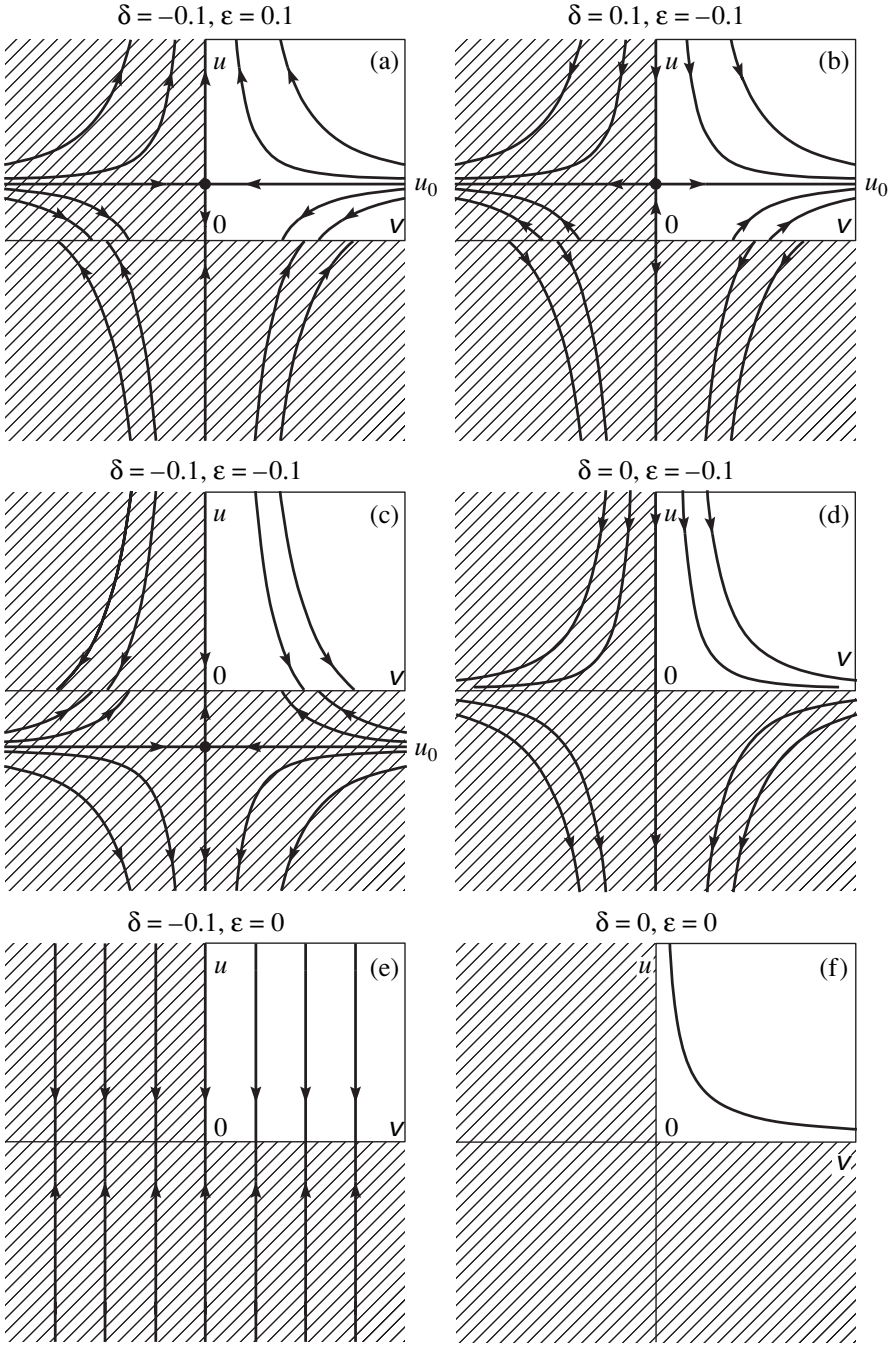


Fig. 1. Phase portraits of the dynamical system for $\mu = 0$ and $\beta = 0$ (cubic GLE).

the functions

$$P_1(u) = \mu a_2 u^2 + 2\epsilon a_1 u + \delta \tag{29}$$

and

$$Q_1(u) = \epsilon + \mu u \tag{30}$$

simultaneously. It is obvious that such equilibria exist only when ϵ and μ have opposite signs. An analysis of the system reveals the following possible types of phase portraits.

I. If $\Delta_1 < 0$, then $P_1(u)$ does not vanish for any δ, ϵ, or μ. There are no nontrivial or degenerate singular points. The only singular line is $u = 0$. Orbits are described by Eq. (23). Pumping, losses, and dispersion cannot balance out.

II. If $\Delta_1 = 0$, then trivial singular points coexist with the saddle-node bifurcation at $(0, u_0)$ (see Fig. 3d

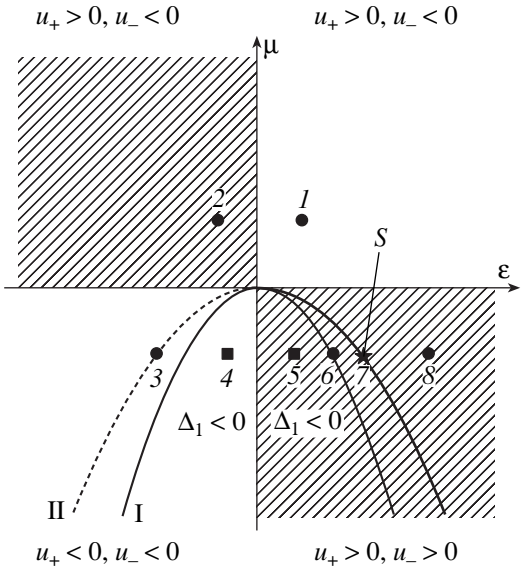


Fig. 2. Bifurcation diagram in the εμ plane for β = 0 and δ < 0: $u_\pm$ denotes the roots of the quadratic trinomial P_1; *I* is the line where a degenerate singular point originates ($\mu = a_1^2\epsilon^2/a_2\delta$); *II* is the line where nontrivial singular points exist ($\mu = 2\epsilon^2/15\delta$); points *1–7* are examples of typical phase-portrait topology; *S* represents a soliton.

below), and there are no singular points of other types. Orbits are described by Eq. (24).

III. If $\Delta_1 > 0$, then the phase portrait is qualitatively different. If one of the roots of the function in (29),

$$u_\pm = -u_1 \pm \Delta_1^{1/2},$$

equals the single root of Eq. (30),

$$u_0 = -\epsilon/\mu,$$

then the phase plane contains a singular line,

$$u = -\epsilon/\mu.$$

Moreover, there also exists a singular point corresponding to the other root. The roots may or may not belong to the domain of admissible *u* and *v*. This depends on the specific values of physical parameters. Orbits are described by Eq. (25). When the roots are not equal, there exist the degenerate singular points $(0, u_+)$ and $(0, u_-)$, where u_+ and u_- are the roots of (29). These points may or may not lie in the upper half-plane. This depends on the signs of the roots, i.e., on the physical parameters contained in the equation. The nontrivial singular points exist if

$$\mu = \frac{2\epsilon^2}{15\delta}, \quad \epsilon\mu < 0. \tag{31}$$

Under the latter condition, the singular line lies in the domain of the phase portrait.

The invariance of system (20), (21) implies that any point $A(\epsilon_0, \mu_0)$ in the εμ plane corresponding to δ < 0 is associated with a point $A'(-\epsilon_0, -\mu_0)$ corresponding to δ > 0. The phase portraits in the neighborhoods of mutually associated points are geometrically identical, differing only by orbit direction. Figure 2 shows the bifurcation diagram for dynamical system (20), (21) with δ < 0. The hatched quadrants may contain nontrivial singular points. The curves $\mu = a_1^2\epsilon^2/a_2\delta$ (where a singular point originates) and $\mu = 2\epsilon^2/15\delta$ (where nontrivial singular points exist) divide the εμ plane into domains of typical phase-portrait topology.

Figure 3 illustrates the phase-portrait topology near the characteristic points of the bifurcation diagram shown in Fig. 2. Since $\Delta_1 > 0$ at points *1* and *2* (in the first and second quadrants), there exist roots of $P_1(u)$ having opposite signs. Of the degenerate saddle points $(0, u_-)$ and $(0, u_+)$, only $(0, u_+)$ lies in the domain of the phase portrait (see Figs. 3a and 3b). The domains above and below the line $u = u_+$ correspond to antidissipative compressive and dissipative compressive pulse dynamics (cf. Fig. 1a), whereas the dynamics at the associated points *1*' and *2*' are antidissipative dispersive and dissipative dispersive, respectively, as in Fig. 1b. In the third quadrant, $u_- < 0$ and $u_+ < 0$ to the left of curve *I* (where $\Delta_1 = 0$); i.e., the domain of the phase portrait does not contain any degenerate singular points. To the right of curve *I*, $\Delta_1 < 0$ and there are no degenerate singular points. The corresponding phase portraits are illustrated by Figs. 3c and 3d, respectively (cf. Fig. 1c).

The transformation of the phase portrait is more complicated when the parameters lie in the fourth quadrant of the bifurcation diagram. Since $\Delta_1 < 0$ to the left of curve *I* (point *5*), the trinomial $P_1(u)$ has no roots and, therefore, there are no degenerate singular points (see Fig. 3e). At point *6* (on curve $\Delta_1 = 0$), there is a saddle-node bifurcation (Fig. 3g) [5]. To the right of the curve, $P_1(u)$ has two positive roots and the saddle-node bifurcation yields a saddle point and a node (see Fig. 3h). At point *7* (on curve *II*), the singular line $u = u_+$ originates. If δ < 0, then every trajectory passing through a point above this line and a point between the lines $u = u_-$ and $u = u_+$ terminates at a point on this line (see Fig. 3j). When the combination of parameters in Eq. (2) corresponds to this last case, we should expect a stable soliton solution to exist.

Note that the singular line is a structurally unstable (nonrobust) set; i.e., it is destroyed by a small change in parameters. This is important because the constants a_0, a_1, a_2, and *b* contain summands proportional to $1/\pi^2$; i.e., solitons seem unlikely to exist in this case at first glance. Figures 3i and 3k show the phase portraits corresponding to the left and right neighborhoods of point *7*, where $(0, u_+)$ is a node and a saddle point, respectively. Both phase portraits are characterized by extremely slow dynamics near the separatrix $u = u_+$ and extremely narrow (numerically irresolvable) gaps between the upper and lower values of *u* in its neighbor-

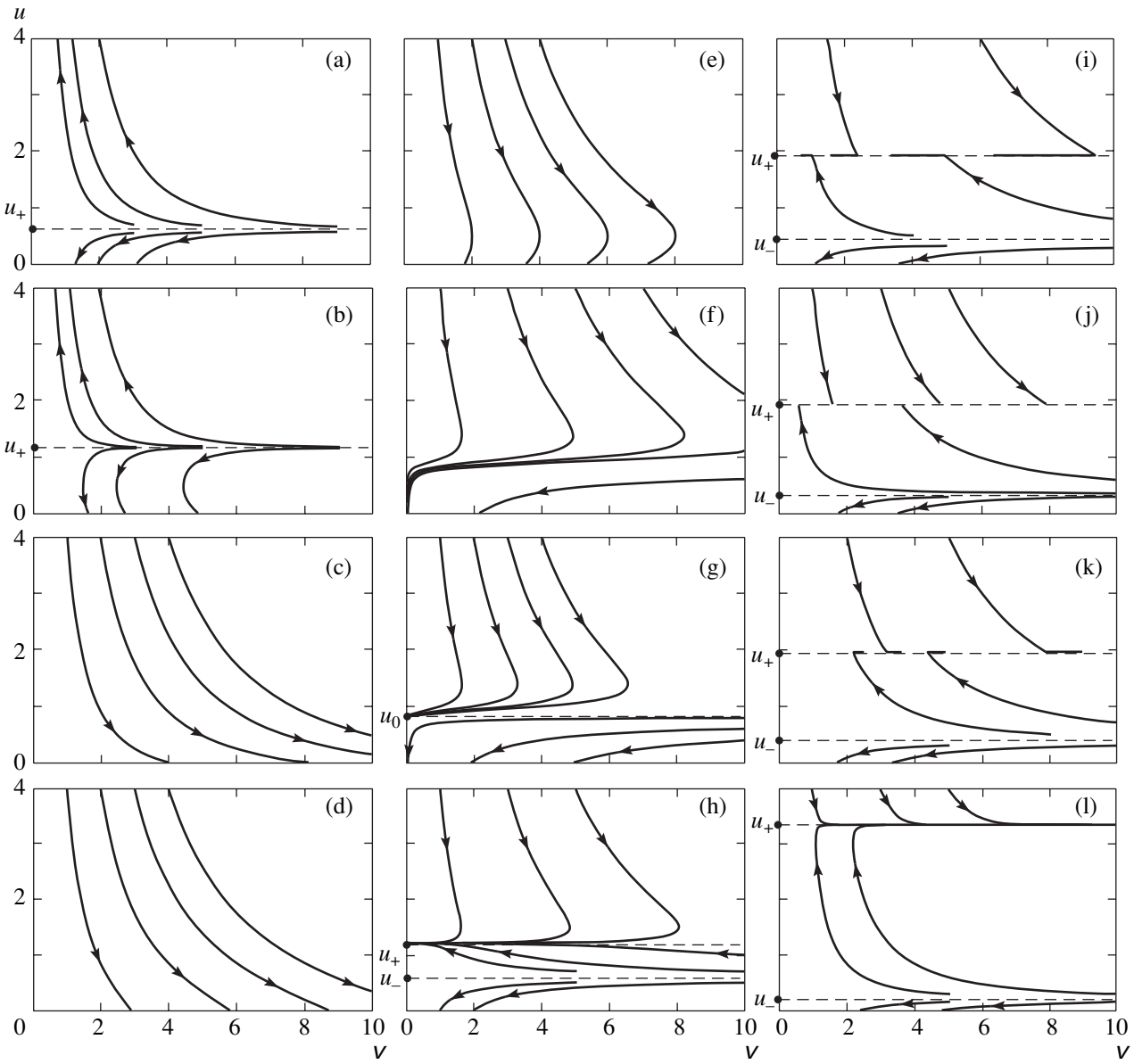


Fig. 3. Phase portraits of dynamical system (20), (21) with $\delta = -0.1$ at the typical points of the bifurcation diagram shown in Fig. 2. The singular line in (k) and the horizontal separatrices associated with the degenerate singular points $(0, u_+)$ and $(0, u_-)$ are represented by dashed lines. Vertical separatrices are ordinate axes. (a) $\mu = 0.2$, $\epsilon = 0.1$ (point *1*); (b) $\mu = 0.2$, $\epsilon = -0.1$ (point *2*); (c) $\mu = -0.2$, $\epsilon = -0.387$ (point *3*); (d) $\mu = -0.2$, $\epsilon = -0.1$ (point *4*); (e) $\mu = -0.2$, $\epsilon = 0.1$ (point *5*); (f) $\mu = -0.2$, $\epsilon = 0.274$ (left neighborhood of point *6*); (g) $\mu = -0.2$, $\epsilon = 0.277$ (point *6*); (h) $\mu = -0.2$, $\epsilon = 0.3$ (right neighborhood of point *6*); (i) $\mu = -0.2$, $\epsilon = 0.380$ (left neighborhood of point *7*); (j) $\mu = -0.2$, $\epsilon = 0.38722$ (point *7*); (k) $\mu = -0.2$, $\epsilon = 0.388$ (right neighborhood of point *7*); (l) $\mu = -0.2$, $\epsilon = 0.6$ (point *8*).

hoods. In fact, one can say that orbits approach some limit values when the parameters δ, ϵ, and μ belong to a certain set rather than obey a strict relation. The behavior of orbits outside this set agrees with that predicted theoretically for the neighborhood of a separatrix (see Fig. 3l).

3.2. Filtered Problem ($\beta \neq 0$)

In this case, the common feature of the phase portrait of system (18), (19) is that the line $v = 0$, including the origin, does not belong to the domain of the phase portrait (since the right-hand side of Eq. (18) contains a term proportional to v^{-1}). However, the equation

$$\frac{du}{dv} = \frac{-\beta a_0 u + uv(\delta + 2\epsilon a_1 u + \mu a_2 u^2)}{v[2\beta b - buv(\epsilon + \mu u)]}$$

has the saddle point $O(v = 0, u = 0)$ with the separatrices $u = 0$ and $v = 0$. Thus, the lines $u = 0$ and $v = 0$ also are limit sets of system (18), (19).

3.2.1. The case of $\mu = 0$

In this case, system (18), (19) reduces to

$$u' = -2\beta a_0 \frac{u}{v} + 2u(\delta + 2\epsilon a_1 u), \quad (32)$$

$$v' = 4\beta b - 2buv\epsilon. \quad (33)$$

Unlike system (20), (21), it has neither singular lines nor degenerate singular points on the axes. When either $\delta > 0$, $\epsilon < 0$, and $\beta < 0$ or $\delta < 0$, $\epsilon > 0$, and $\beta > 0$, there exists a singular point $(-2\delta/\epsilon, -\beta/\delta)$ in the first quadrant. When the combination of signs of these parameters is different, the singular point lies outside the domain of the phase portrait and the dynamics of a localized distribution should be classified as either dissipative dispersive (the amplitude tends to zero) or antidissipative compressive (the distribution tends to collapse).

Figure 4 illustrates the dynamics of a system with a saddle point. The quadrant is divided by separatrices into four regions. When $\delta > 0$, the system exhibits antidissipative dynamics in regions *I* (between the coordinate axes and the separatrices) and *II* (between the separatrices above the singular point) and dispersive dynamics in regions *III* (between the separatrices below the singular point) and *IV* (on the right and above the separatrices).

Note that relatively wide and narrow pulses having equal amplitudes exceeding a certain threshold value exhibit dispersive and antidissipative dynamics, respectively. Pulses with amplitudes below the threshold are always characterized by antidispersive dynamics. The existence of a gap between the lower branch of a separatrix and the axis $u = 0$ means antidispersive instability of the background. When $\delta < 0$, the separatrices and orbits have opposite directions.

As above, the actual dynamics depend on the relative position of the NLS soliton curve and the saddle separatrix, which is determined by the values of δ, β, and ϵ.

3.2.2. The case of $\mu \neq 0$

When dynamical system (18), (19) is a complete one, the singular points are the roots of the expressions

$$P_2 = -2\beta a_0 \frac{u}{v} + 2u(\delta + 2\epsilon a_1 u + \mu a_2 u^2), \quad (34)$$

$$Q_2 = 4\beta b - 2buv(\epsilon + \mu u) \quad (35)$$

on the right-hand sides. Solving (35) for v and substituting the result into (34), we obtain

$$11\mu u^2 + 15\epsilon u + 30\delta = 0, \quad (36)$$

which yields the following expressions for the coordinates of singular points:

$$u_\pm = \frac{-15\epsilon \pm \Delta_2^{1/2}}{22\mu}, \quad v_\pm = \frac{2\beta}{u_\pm(\epsilon + \mu u_\pm)}, \quad (37)$$

where

$$\Delta_2 = (15\epsilon)^2 - 1320\mu\delta.$$

Therefore, singular points exist if

$$u_\pm \neq 0, \quad u_\pm \neq -\epsilon/\mu, \quad \Delta_2 > 0.$$

These conditions should be supplemented with

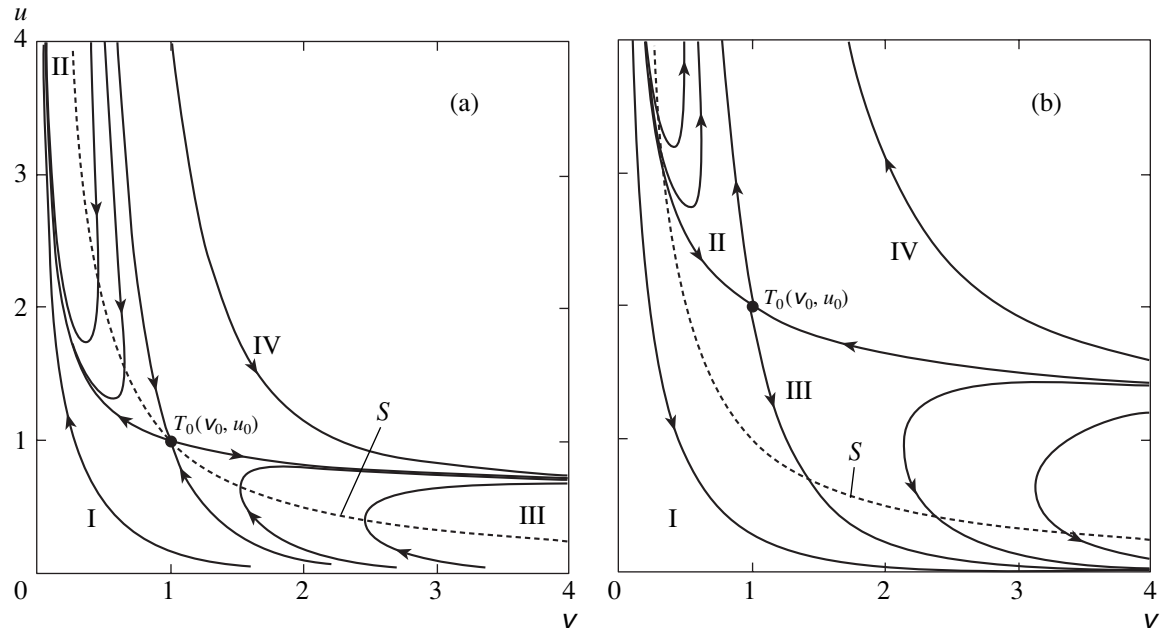


Fig. 4. Phase portraits of the dynamical system for $\beta \neq 0$ and $\mu = 0$: (a) $\delta = 0.1$, $\epsilon = -0.2$, $\beta = -0.1$; (b) $\delta = -0.1$, $\epsilon = 0.1$, $\beta = 0.1$. Dashed curve *S* is the NLS soliton curve (initial conditions).

$$u_{\pm} > 0, \quad v_{\pm} > 0.$$

By the condition $\Delta_2 > 0$, the elliptic cone $15\epsilon^2 - 88\delta\mu = 0$ divides the three-dimensional space (δ, ϵ, μ) into an exterior set, which contains singular points, and an interior one, which does not. For any particular $\delta \neq 0$, the distributions of singular points in the $\epsilon\mu$ plane (intersecting the cone) that correspond to all possible combinations of signs of δ and β can be represented by the four diagrams shown in Fig. 5. Here, curve *I* (parabola $\mu = 15\epsilon^2/88\delta$) is the intersection of the conic surface and the plane δ = const, and curve *II* (parabola $\mu = 2\epsilon^2/15\delta$) is the locus of points where either v_+ or v_- changes sign. The equation for this curve is obtained by setting to zero the denominator of the expression for $v_{\pm}$ in (37). Note that this equation is identical to the condition for existence of a singular line in the case of $\beta = 0$ (see Section 3.1). If $\delta\mu < 0$, then the system has two singular points, T_+ and T_-, which lie between parabolas *I* and *II*.

It can readily be shown (see Appendix B) that the singular points of system (18), (19) that exist for any combination of parameters are saddle points. Moreover, the gap between parabolas *I* and *II* contains a node, which is stable if $\delta < 0$ and $\beta > 0$ and unstable if $\delta > 0$ and $\beta < 0$. By the index theorem, these singular points are incompatible with limit cycles. In terms of the original problem formulation, this means that there are no monophase *t*-periodic (unchirped) solutions to GLE bifurcating from the NLS soliton.

Figure 6 shows the phase portraits with two singular points corresponding to two specific sets of parameter

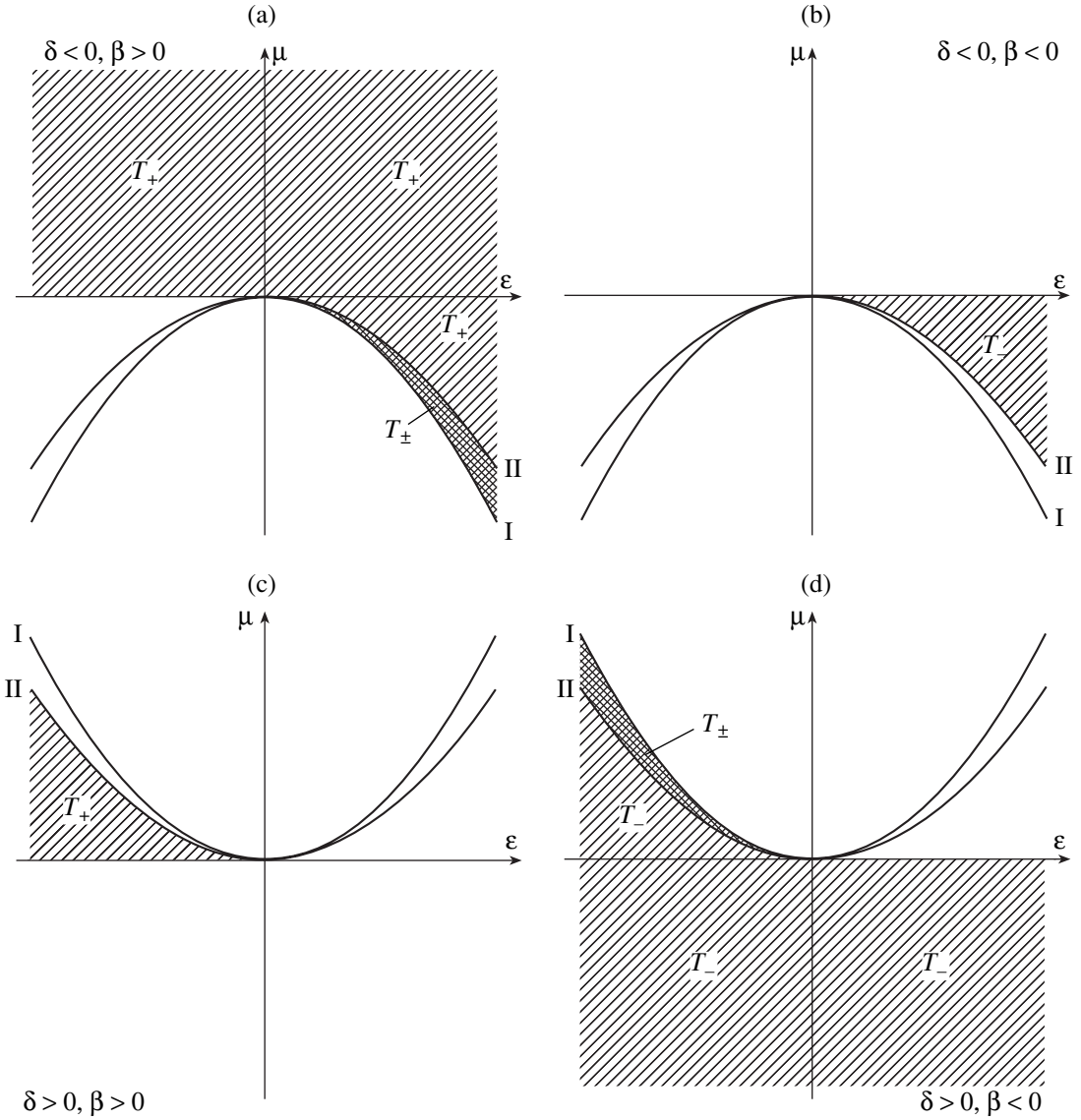


Fig. 5. Bifurcation diagrams for $\beta \neq 0$ for different signs of β and δ: hatched areas contain saddle points T_+ or T_-; the gap between parabolas *I* and *II* contains two singular points; unhatched areas contain no singular points.

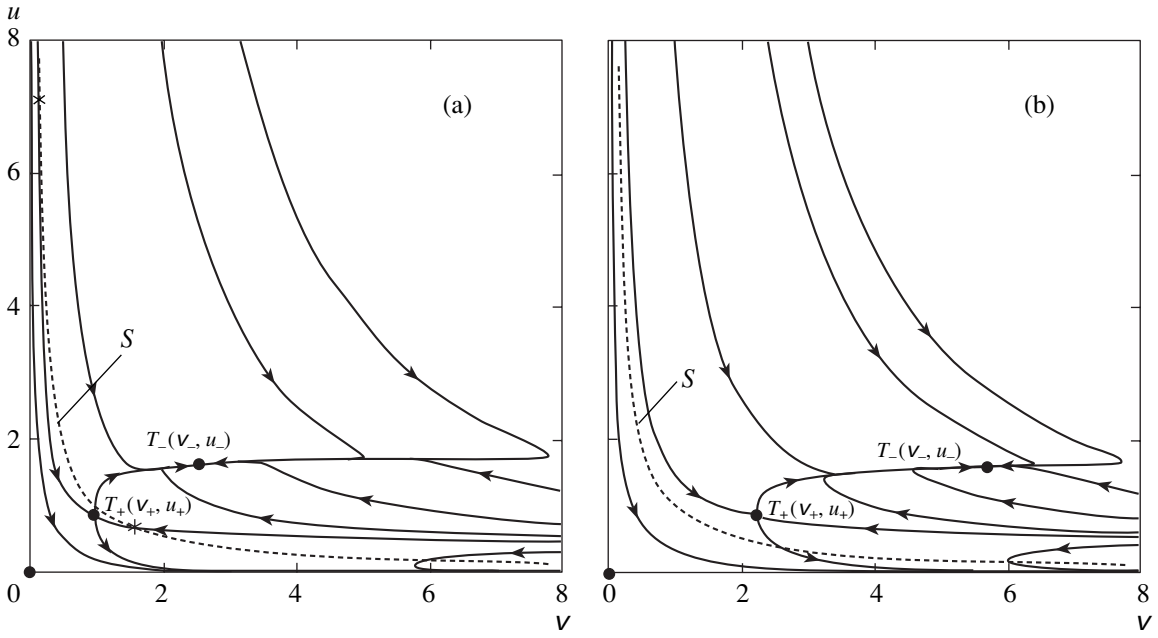


Fig. 6. Phase portraits of the system with a saddle point and a node, T_+ and T_-, for $\beta \neq 0$: dashed curve S is the NLS soliton curve; asterisks represent the limit values of initial conditions that lead to the formation of a soliton. (a) $\delta = -0.1$, $\epsilon = 0.36$, $\mu = -0.2$, $\beta = 0.05$; (b) $\delta = -0.1$, $\epsilon = 0.36$, $\mu = -0.2$, $\beta = 0.18$.

values (cf. Fig. 3i, with a saddle point and a stable node on the line $v = 0$). Note that the saddle-node point corresponding to a specific μ and ϵ lying on curve I splits into a saddle point and a node as ϵ increases. The stable node $T(v_-, u_-)$ corresponds to a stable soliton solution to the original problem (see Fig. 11a). Its basin of attraction contains only trajectories that originate from the NLS soliton curve (dashed curve S in Fig. 6a) and lie above the saddle separatrix T_+. The relative positions of the NLS soliton curve and point T_+ depend not only on the values of δ, ϵ, and μ, but also on β. When β (i.e., the coordinate v_+) is varied while the pumping/loss parameters δ, ϵ, and μ are held constant, the curve S may lie outside the basin of attraction of the node. Since the equation for the NLS soliton curve is

$$uv = 1,$$

the curve S lies (at least, partially) in the basin of attraction of the node if

$$u_+ < 1/v_+,$$

which implies that the parameters satisfy the relation

$$\beta < \frac{7\epsilon + \Delta_2^{1/2}}{44}. \qquad (38)$$

When $\delta\mu \neq 0$, the upper boundary of the set of admissible ϵ and β is the upper half of the right-hand branch of the hyperbola

$$\frac{(\epsilon')^2}{a^2} - \frac{(\beta')^2}{b^2} = 1,$$

where

$$a^2 = 6\delta\mu, \quad b^2 = 2\delta\mu/3,$$

and the canonical coordinate system (ϵ', β') makes the angle

$$\phi = \arcsin 1/\sqrt{50}$$

with (ϵ, β). The left and right boundaries of the set are

$$\epsilon = \epsilon_I, \quad \epsilon = \epsilon_{II},$$

where

$$\epsilon_I = \sqrt{\frac{88}{15}\delta\mu}, \quad \epsilon_{II} = \sqrt{\frac{15}{2}\delta\mu}.$$

Both the angle between the asymptotes of the hyperbola,

$$\alpha = 2a/b = 2\arctan 3,$$

and the angle ϕ are independent of the parameters, unlike the apex coordinates

$$x_0 = \sqrt{\frac{7}{75}\delta\mu}, \quad y_0 = \sqrt{\frac{1}{75}\delta\mu}.$$

When $\delta\mu = 0$, the hyperbola collapses into the line $\beta = (1/2)\epsilon$ (one of the asymptotes) and the domain in ϵ

shrinks to a point (origin). Figure 7 shows the basin of attraction of the node T_- in the parameter plane (ϵ, β) for $\delta = -0.1$ and $\mu = -0.2$. Here,

$$\epsilon_I = 0.34254\ldots, \quad \epsilon_{II} = 0.38730\ldots, \quad \beta_{max} = 0.1229.$$

It follows from the analysis presented above that β_{max} cannot reach the asymptote $\beta = (1/2)\epsilon$ for any set of parameter values. In terms of the original problem formulation, this means that, even though there exists a monophase stationary solution for $\beta > \beta_{max}$, it cannot be reached if the initial conditions correspond to an NLS soliton. This is also true for the exact monophase solution reported in [6] (see Fig. 6b). Since the NLS soliton curve S lies below the separatrices of the saddle point T_+, the amplitude of a localized solution tends to zero for any set of initial conditions.

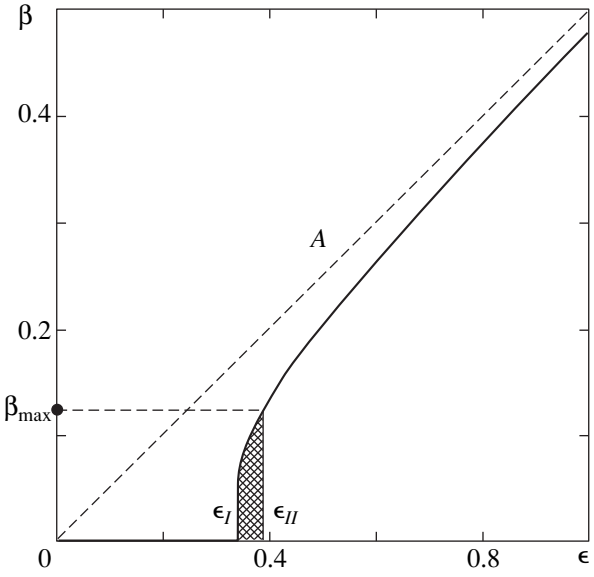


Fig. 7. Basin of attraction of a stable soliton solution in the parameter plane (ϵ, β) (hatched): ϵ_I and ϵ_{II} are, respectively, the extreme left and right values of ϵ for which a stable node exists; the parameters δ and μ are held constant.

4. NUMERICAL SOLUTIONS OF THE GINZBURG–LANDAU EQUATION

To verify the results of our qualitative analysis of the dynamical system, we computed the evolution of the field initially corresponding to an exact solution to the NLS equation with cubic nonlinearity,

$$\psi(t = 0, x) = \frac{q}{\cosh[q(x - x_c)]},$$

where q = const and $x_c = (x_N - x_0)/2$ is the center-of-mass coordinate. Equation (2) was solved by a finite-difference method. The numerical technique is described in Appendix B. The computations were performed for the following parameter values:

$$\delta = -0.1, \quad \mu = -0.1, -0.2, -0.5, \quad \nu = 0.$$

The results obtained are presented below.

Case 1: $\beta = 0$, $\mu = 0$, $\epsilon \neq 0$, $\delta \neq 0$ $\nu = 0$. As shown above, dynamical system (20), (21) has only the singular line $u = 0$ and a saddle point $(0, u_0)$, where $u_0 = -\delta/2\epsilon a_1$. This result is corroborated by numerical solution of the original GLE. When $\delta < 0$ and $q < u_0$, the amplitude of a localized initial disturbance tends to zero with increasing t, i.e., the pulse decays (damped by predominant dissipation). When $q > u_0$, the pulse amplitude rapidly increases over the run time (due to predominant pumping, cf. Fig. 1a).

When $\delta > 0$, dynamical system (20), (21) exhibits dispersive dynamics: the pulse width grows indefinitely, while its amplitude is bounded (cf. Fig. 1b). Figure 8 shows the numerical results obtained by solving the GLE for the parameter values corresponding to Fig. 1b. The trends illustrated by these figures obviously correlate: the pulse width increases while the peak amplitude does not exceed a certain threshold value. It is obvious that the complicated structure of the excited region cannot be resolved in a phase portrait obtained by numerical approximation. Figure 8b shows a "photographic" image of $|\psi(x, t)|$, where the grayscale value is inversely proportional to the amplitude of the function. The image clearly demonstrates the regular spatiotemporal structure of the solution. Ripples at large t are due to boundary effects. Figure 9 shows the amplitude and phase of the solution at $t = 30$.

Case 2: $\beta = 0$, $\mu \neq 0$, $\epsilon \neq 0$, $\delta \neq 0$, $\nu = 0$. When $\delta > 0$, the numerical solution of GLE are fully consistent with our qualitative analysis of dynamical system (20), (21). The dynamics corresponding to $\delta < 0$ and $\mu > 0$ (i.e., first and second quadrants of the parameter plane (ϵ, μ), see Fig. 2) also agree with the qualitative theory, and so do the dynamics associated with parameter values in the third quadrant. For $\epsilon = \sqrt{a_2\delta\mu}/a_1$ in the fourth quadrant, we observed a localized soliton solution that did not break down over the run time, in complete agreement with the qualitative theory. Figure 10 shows the amplitude and phase of this soliton. When the amplitude is large, the almost linear behavior of the phase function is close to unchirped trial function (8). The peak amplitude of the soliton is approximately 1.39 (cf. $u_+(=A^2) \approx 1.93$ for the dynamical system). Note that solitons of this kind are obtained by solving GLE for ϵ in a relatively wide interval (from 0.31 to 1.79 when $\delta = -0.1$ and $\mu = -0.2$) and the squared soliton amplitude is close to u_+ (saddle-point coordinate). When both δ and μ are held constant, the soliton amplitude and width increase with ϵ. Note that the characteristics of these solitons correlate with the phase portrait described above.

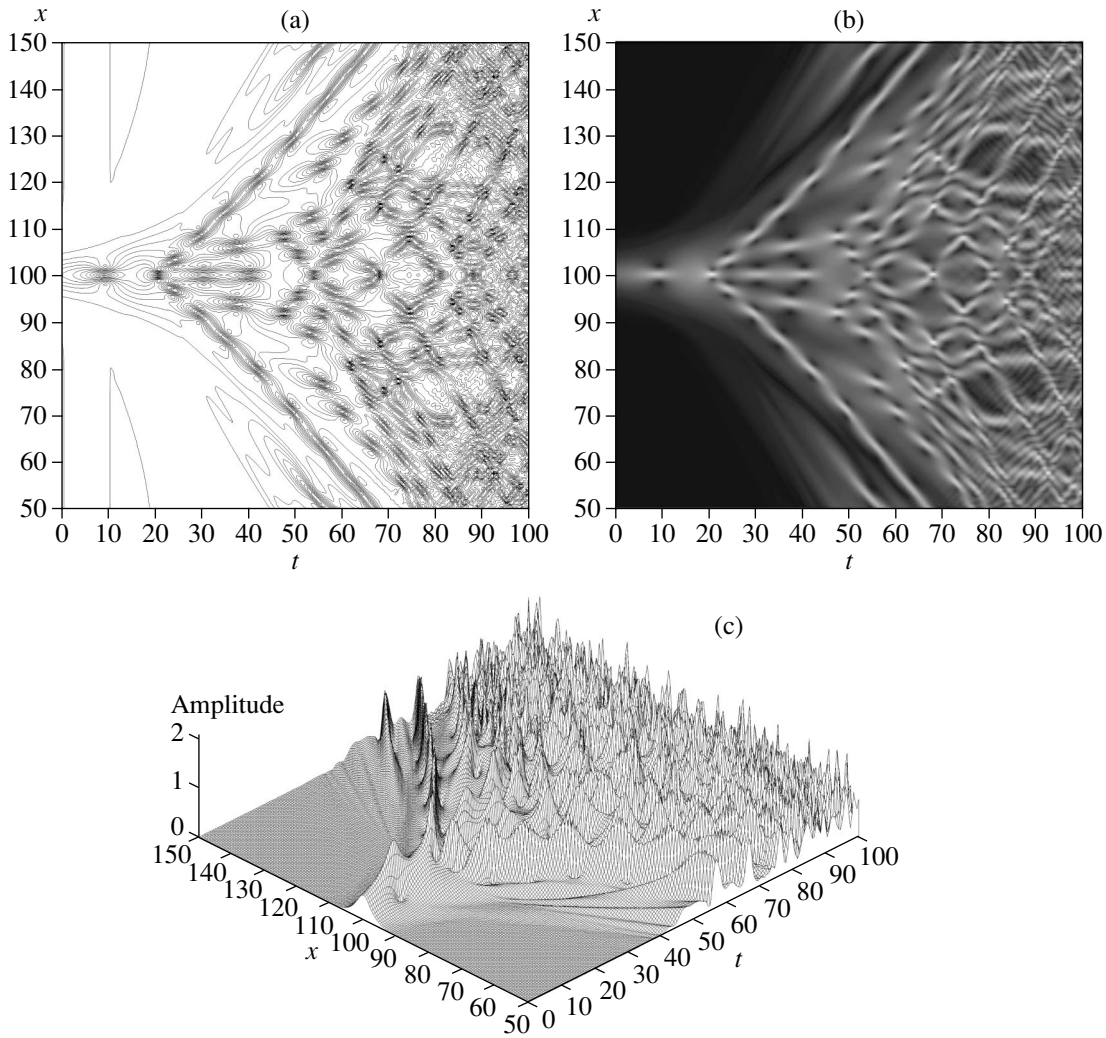


Fig. 8. Antidissipative dispersive dynamics for $\delta = 0.1$, $\epsilon_I = -0.1$, $\mu = 0$, $\beta = 0$, and $\nu = 0$ for the initial amplitude $|\psi(0, x_c)| = 0.5$ at the center of symmetry: (a) contour map of $|\psi(t, x)|$; (b) "photographic" image with grayscale value is inversely proportional to $|\psi(t, x)|$; (c) three-dimensional map of $|\psi(t, x)|$.

Beyond $\epsilon = 1.79$, the dynamics are as predicted by the qualitative theory (dissipative and dispersive for small and large amplitudes, respectively, similar to Fig. 8).

According to the classification proposed in [6], the computed solitons are fixed-amplitude solitons, because their amplitudes are uniquely determined by the parameters of the original equation. However, the domain of existence of these solitons is different from the domain of fixed-amplitude solitons obtained in [6] as analytical solutions to the quintic GLE. When $\beta = 0$ and $\nu = 0$, the latter solitons exist for $\epsilon = \pm 3\sqrt{3}$ and an arbitrary μ. One obvious conclusion is that neither fixed-amplitude solitons found in [6] nor solitons with arbitrary amplitudes can be obtained under the assumptions made about the form of trial function (8). This is an expected result because the solitons are chirped (their phase functions depend on x).

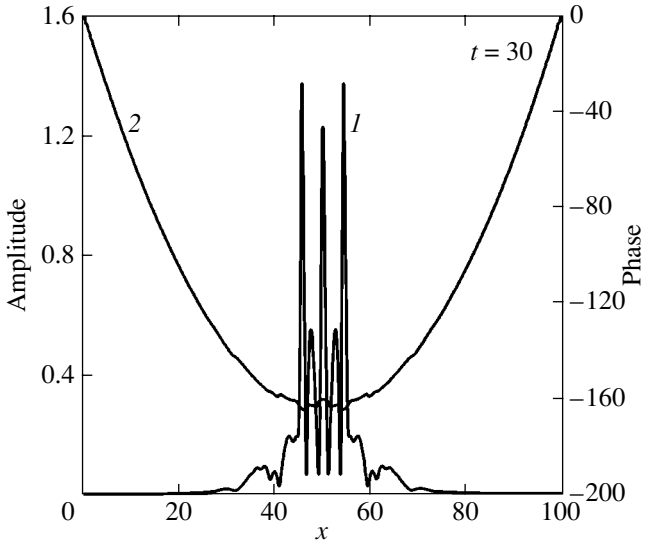


Fig. 9. Amplitude (*1*) and phase (*2*) of $\psi(t, x)$ at $t = 30$ for $\delta = 0.1$, $\epsilon = -0.1$, $\mu = 0$, and $\beta = 0$.

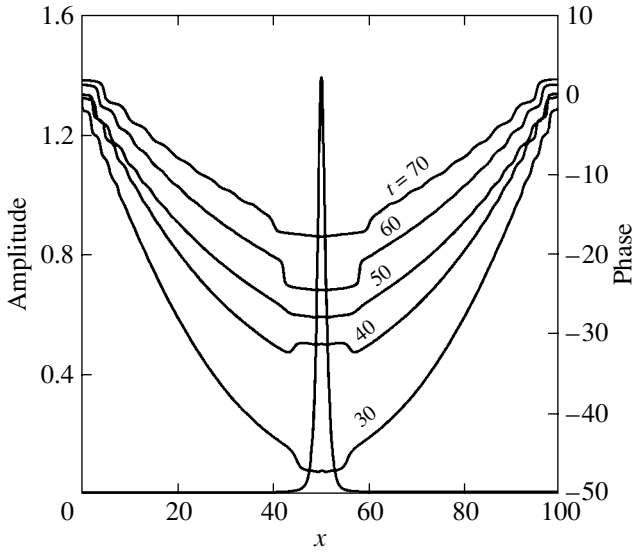


Fig. 10. Amplitude and phase of a soliton at several instants for $\beta = 0$, $\delta = -0.1$, $\epsilon = 0.38722$, $\mu = -0.2$, and $\nu = 0$.

Case 3: $\beta \neq 0$, $\mu = 0$, $\epsilon \neq 0$, $\delta \neq 0$, $\nu = 0$. In this case, the results of direct numerical solution of GLE agree with our qualitative analysis as well: no new monophase localized solutions to GLE bifurcate from the NLS soliton. The dynamics observed when $\delta < 0$ are either dissipative or antidissipative. When $\delta > 0$, the dynamics are either dispersive or antidissipative, depending on β and initial amplitude according to the relative positions of the NLS soliton curve and the saddle point.

Case 4: $\beta \neq 0$, $\mu \neq 0$, $\epsilon \neq 0$, $\delta \neq 0$, $\nu = 0$. In this case, the conclusions based on our qualitative analysis are also in good agreement with the results obtained by numerical solution of GLE. Soliton solutions were obtained in the interval between ϵ_I and ϵ_{II} for $\delta = -0.1$ and $\mu = -0.1$, -0.2, and -0.5. Figure 11a shows the amplitude and phase of the soliton obtained by solving GLE with $\delta = -0.1$, $\epsilon = 0.36$, $\mu = -0.2$, and $\beta = 0.08$. These parameter values correspond to a point between parabolas *I* and *II*, where a stable node and a saddle point exist according to the qualitative theory. However (as in Case 2, when $\beta = 0$ and $\mu \neq 0$), the right boundary of the domain of existence of the soliton solution is located at a considerable distance to the right of the value of ϵ_{II} predicted by the qualitative theory. In our computations, this value reached 1 for $\mu = -0.2$ and 1.5 for $\mu = -0.5$. It should be noted that the phase diagrams corresponding to $\epsilon > \epsilon_{II}$ in the neighborhood of curve *II* are similar to those corresponding to $\beta = 0$ in that the orbits approaching the saddle separatrix are characterized by slow dynamics. However, the results obtained by direct numerical solution of GLE for $\epsilon_{II} < \epsilon < \epsilon_{max}$ clearly demonstrate that the phase function substantially deviates from linear behavior even in the central region (see Fig. 11b). Beyond ϵ_{max}, our computations revealed complex irregular dynamics involving expansion of the excited region, with bounded amplitude. The example shown in Fig. 12 corresponds to $\delta = -0.1$, $\epsilon = 1.01$, $\mu = -0.2$, and $\beta = 0.015$. Dispersive instability develops rapidly, and it may be difficult to isolate the effects due to the computational-domain boundary. However, the initial stage of the instability can be referred to the dispersive type that is characteristic of a dynamical system with a saddle point.

With further increase in ϵ, the dynamics of the localized solution goes through a number of intricate changes. Ultimately (for $\epsilon > 1.8$), we obtained two

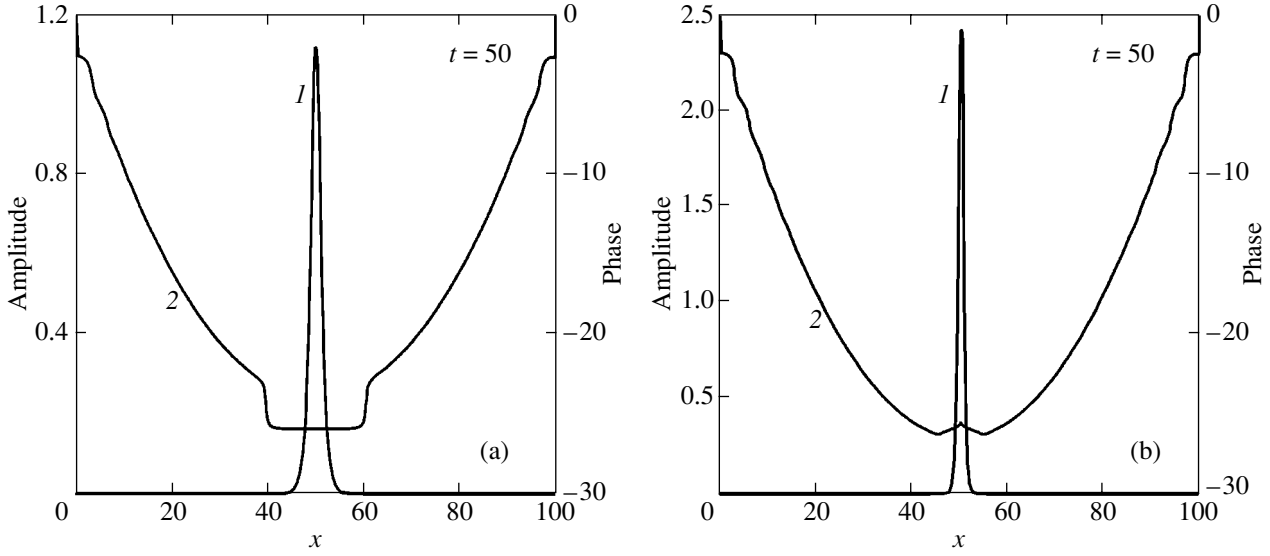


Fig. 11. Amplitude and phase of a stable soliton for $\beta \neq 0$: (a) $\delta = -0.1$, $\epsilon = 0.36$, $\mu = -0.2$, $\beta = 0.08$ (below curve *II*); (b) $\delta = -0.1$, $\epsilon = 0.99$, $\mu = -0.2$, $\beta = 0.08$ (above curve *II*).

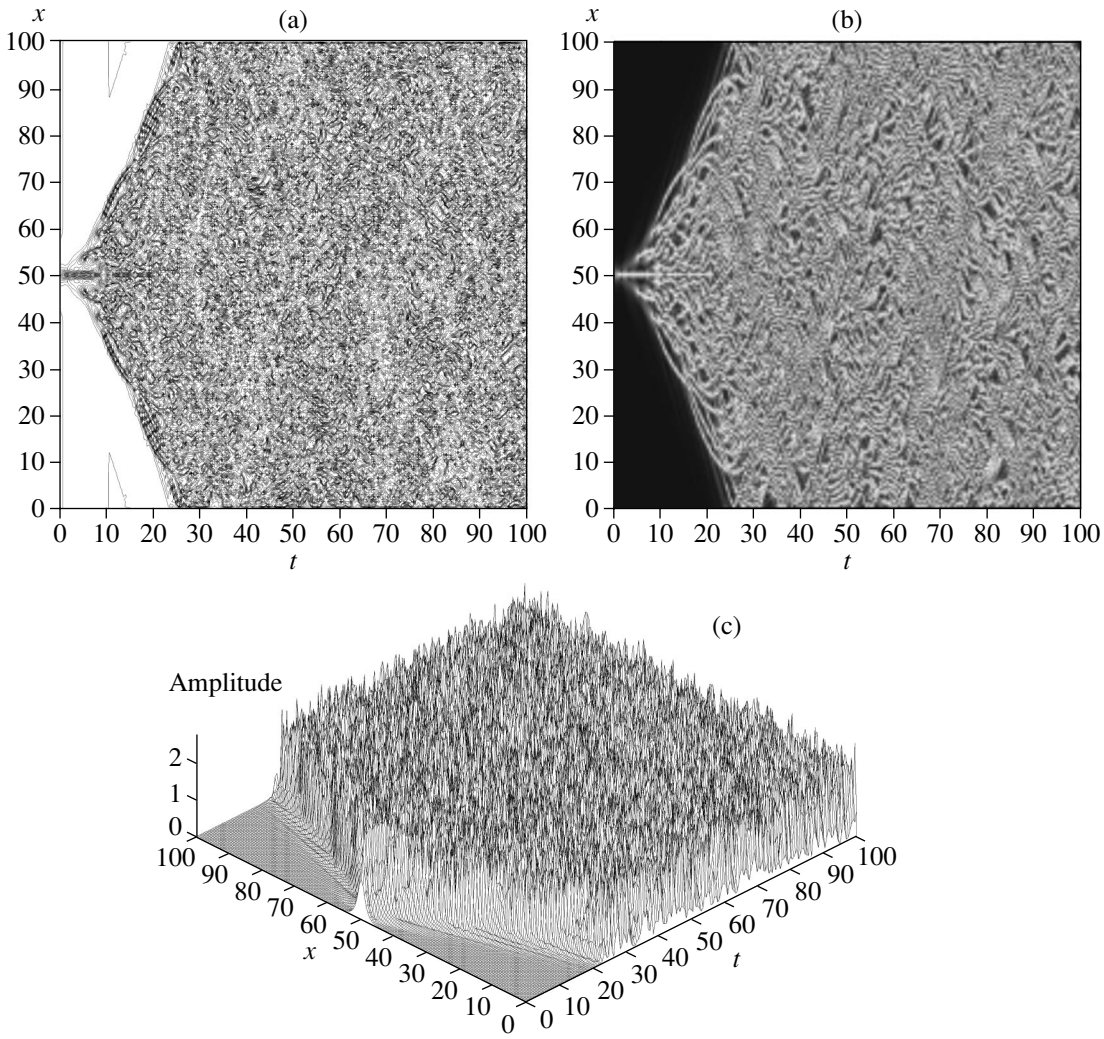


Fig. 12. Chaotic dynamics of a solution to GLE for $\delta = -0.1$, $\epsilon = 1.01$, $\mu = -0.2$, and $\beta = 0.015$: (a) contour map of $|\psi(x, t)|$; (b) "photographic" image; (c) three-dimensional map.

diverging fronts, i.e., dynamics that can be classified as dispersive.

5. CONCLUSIONS

1. The generalized moment method is used to construct a dynamical system describing the dynamics of amplitude, width, center-of-mass coordinate, phase function, and wave number of solutions to the Ginzburg–Landau equation.

2. A qualitative analysis of the dynamical system is performed for symmetric field distributions. Singular points and singular lines are found. It is shown that the dynamical system does not have periodic solutions. Bifurcation diagrams are presented, and domains where stable stationary soliton solutions exist are determined. Typical phase portraits are depicted.

3. Good agreement is established between results of the qualitative analysis and solutions obtained by direct computation of GLE for the same values of physical parameters.

4. The numerical analysis has revealed a new branch of soliton solutions of the original GLE that are different from the known monophase fixed-amplitude solitons found analytically in [6].

APPENDIX A

Analysis of Stability of Singular Points in the First Approximation

The characteristic equation for dynamical system (18), (19) is

$$\lambda^2 - C_\pm\lambda + D_\pm = 0, \qquad (39)$$

where

$$C_\pm = \left(\frac{\partial P_2}{\partial u} + \frac{\partial Q_2}{\partial v}\right)\Big|_{T_\pm(u_\pm, v_\pm)},$$

$$D_\pm = \left(\frac{\partial P_2}{\partial u}\frac{\partial Q_2}{\partial v} - \frac{\partial P_2}{\partial v}\frac{\partial Q_2}{\partial u}\right)\Big|_{T_\pm(u_\pm, v_\pm)}.$$

Calculating the derivatives of the expressions for P_2 and Q_2 given by (34) and (35) and using the equations

$$11\mu u_\pm^2 + 15\epsilon u_\pm + 30\delta = 0, \quad v_\pm = \frac{2\beta}{u_\pm(\epsilon + \mu u_\pm)},$$

we obtain

$$C_\pm = -\frac{60}{11}(d\epsilon u_\pm + 16\delta), \tag{40}$$

$$D_\pm = -\frac{4b}{11\mu^2} \times [(253\delta\mu - 30\epsilon^2)\epsilon u_\pm + 60(11\delta\mu - \epsilon^2)\delta)]. \tag{41}$$

To determine the sign of the discriminant $\Omega = C_\pm^2 - 4D_\pm$ of Eq. (39), we substitute $C_\pm$ and $D_\pm$ and isolate the quadratic part of (36) equal to zero. The result is

$$\Omega = \frac{16}{11^3\mu}[G\epsilon u_\pm + H\delta], \tag{42}$$

where

$$G = 11d_1\delta\mu - 15d_2\epsilon^2, \quad H = 11d_0\delta\mu - 30d_2\epsilon^2,$$

$$d_0 = 15^2 \times 16^2 + 60 \times 11b,$$

$$d_1 = 15^2 \times 32d + 23 \times 11b,$$

$$d_2 = 15^2d^2 + 2 \times 11b, \quad d = \frac{13}{3} + \frac{11}{\pi^2}, \quad b = \frac{4}{\pi^2}.$$

To determine if Ω can vanish, we substitute $u_\pm$ into (42) and get rid of irrational terms. As a result, we find that the discriminant vanishes if

$$d_0^2 + z(2d_1^2 - 4d_0d_2 - d_0d_1) + z^2(4d_2^2 + d_0d_1 - 2d_1d_2) = 0, \tag{43}$$

where $z = 15\epsilon^2/11\delta\mu$. The discriminant of the last equation is negative. Since the coefficient of z^2 is positive, the quadratic trinomial on the left-hand side can only be positive. Therefore, the discriminant Ω of characteristic equation (39), being a continuous function, does not change its sign at any admissible value of δ, ϵ, or μ; i.e., it is sufficient to determine the sign at an arbitrary point. Finding that the discriminant is positive, we conclude that the Lyapunov characteristic numbers λ associated with the singular points $T_\pm$ are real. Therefore, the singular points of dynamical system (18), (19) can be only saddles or nodes.

To determine the signs of λ, we find the sign of the term $D_\pm$ in (39). Omitting some algebra, we represent it as

$$D_\pm = \frac{4b\beta}{v_\pm}(\epsilon u_\pm + 4\delta). \tag{44}$$

Since b is a positive constant and $v_\pm$ is strictly positive, the sign of $D_\pm$ corresponding to a particular β is determined by the sign of the expression in brackets. Substituting the expression for $u_\pm$, we find that the sign of $D_\pm$ reverses on the surface

$$15\epsilon^2 - 88\delta\mu = 0,$$

which separates the domain that contains equilibria from the domain that does not (see Section 3.2). Thus, $D_\pm$ has a constant sign in the entire domain of existence of the roots $u_\pm$. Again, it can be determined at an arbitrary point in the domain of admissible parameters. The sign is minus everywhere on the entire $\epsilon\mu$ plane except for a narrow gap between parabolas *I* and *II* (see Fig. 5), where a positive $D_\pm$ also exists. This implies that the domain of existence contains saddle points and there exists a node in the gap between the conic surfaces when $\delta\mu < 0$. Note that $T_+(v_+, u_+)$ is a saddle point ($D_+ < 0$) and $T_-(v_-, u_-)$ is a node ($D_- > 0$) if $\delta < 0$ and $\beta > 0$. When $\delta > 0$ and $\beta < 0$, $T_-(v_-, u_-)$ is a saddle point ($D_- < 0$) and $T_+(v_+, u_+)$ is a node ($D_+ > 0$).

The type of stability is determined by the sign of $C_\pm$ as given by (40):

$$-\frac{30}{11^2\mu}[d\epsilon(-15\epsilon \pm \Delta_2^{1/2}) + 16\delta\mu] \vee 0.$$

By simple transformations, its analysis is reduced to finding the sign of

$$15d\epsilon^2(4 - 11d) - 32\delta\mu.$$

When $\delta\mu > 0$, this expression is obviously negative. Therefore, C_- is negative if $\mu < 0$, $\delta < 0$, and $\beta > 0$, and C_+ is positive if $\mu > 0$, $\delta > 0$, and $\beta < 0$. This means that the node is stable in the former case and unstable in the latter.

As noted above, $D_\pm = 0$ on the bifurcation surface *I*, where equilibria originate; i.e., one of the eigenvalues λ

is zero. It can be shown that the degenerate equilibrium corresponds to a saddle-node bifurcation in this case.

APPENDIX B

Numerical Algorithm for Solving the Ginzburg–Landau Equation

We rewrite GLE as

$$\psi_t = s_0\psi_{xx} + (s_1 + s_2|\psi|^2 + s_3|\psi|^4)\psi, \qquad (45)$$

where ψ is a complex-valued function; x and t are independent variables; and s_0, s_1, s_2, s_3 are complex parameters related to those in Eq. (2) as

$$s_0 = \beta + i\sigma, \quad s_1 = \delta, \quad s_2 = \epsilon + i\eta, \quad s_3 = \mu + i\nu.$$

The problem was computed by using a four-point implicit finite-difference scheme on the uniform grid

$$\Omega_{h_t, h_x} = \{(t_j, x_i)|t_j = jh_t, j \geq 0; x_i = ih_x, 0 \leq i \leq N\},$$

where h_t is the step in t, h_x is the step in x, j denotes a time layer, and i is the index of a point on the layer.

The scheme was constructed so as to take into account the fact that the equation is quasilinear. Successive approximations were used in proceeding from layer to layer. Equation (45) was approximated by the finite-difference equation

$$\frac{(\psi_i^{j+1})_{k+1} - (\psi_i^j)_k}{h_t}$$

$$= s_0\frac{(\psi_{i-1}^{j+1})_{k+1} - 2(\psi_i^{j+1})_{k+1} + (\psi_{i+1}^{j+1})_{k+1}}{h_x^2} + (S_i^{j+1})_k,$$

where

$$(S_i^{j+1})_k = (s_1 + s_2|\psi_i^{j+1}|_k^2 + s_3|\psi_i^{j+1}|_k^4)(\psi_i^{j+1})_k,$$

k is the index of an iteration step in computing the solution on the $(j+1)$th layer ($0 \leq k \leq K$), and $(\psi_i^{j+1})_0$ is set equal to $(\psi_i^j)_K$.

Since the original statement of the problem involves boundary conditions set at infinity, approximation on the boundaries of the computational domain presents a certain difficulty. The standard way of dealing with it would be to use the asymptotic form of the solution at the boundary points. However, this approach could not be applied since the solution is not unique. In preliminary computations, we used "hard," "soft," and periodic boundary conditions. The results obtained for a sufficiently large computational domain differed insignificantly. In this paper, we present the results obtained with the use of periodic boundary conditions, as in [7]. The resulting system of linear algebraic equations is

$$A_0(\psi_N^{j+1})_{k+1} - C_0(\psi_0^{j+1})_{k+1} + B_0(\psi_1^{j+1})_{k+1} + F_0 = 0,$$

$$A_i(\psi_{i-1}^{j+1})_{k+1} - C_i(\psi_i^{j+1})_{k+1} + B_i(\psi_{i+1}^{j+1})_{k+1} + F_i = 0, \quad 1 \leq i \leq N-1,$$

$$(\psi_N^{j+1})_{k+1} = (\psi_0^{j+1})_{k+1},$$

where

$$A_i = \frac{s_0}{h_x^2}, \quad C_i = \frac{2s_0}{h_x^2} + \frac{1}{h_t},$$

$$B_i = \frac{s_0}{h_x^2}, \quad F_i = \frac{1}{h_t}(\psi_i^j)_K + (S_i^{j+1})_k.$$

The system was solved by using a cyclic tridiagonal algorithm [8]. The scheme employed in the computations is accurate to $O(h_t + h_x^2)$. The program was debugged by computing well-known exact solutions to the GLE: NLS solitons ($R \equiv 0$) and GLE solitons of constant and arbitrary amplitude [2]. The algorithm has proved effective and reliable. Using this numerical technique, we reproduced the results reported in [3], where a split-step method was applied in combination with fast Fourier transform. The results presented in this paper were obtained for $h_t = 0.001$ and $h_x = 0.05$. The number of iteration steps performed in proceeding from layer to layer was $K = 3$.

REFERENCES

1. A. D. Polyanin and V. F. Zaitsev, *Handbook on Nonlinear Partial Differential Equations* (Fizmatlit, Moscow, 2002; Chapman & Hall/CRC, Washington, 2003).
2. N. N. Akhmediev and A. Ankevicz, *Solitons: Nonlinear Pulses and Beams* (Chapman & Hall, London, 1997; Fizmatlit, Moscow, 2003).
3. J. M. Soto-Crespo, N. Akhmediev, and A. Ankiewicz, Phys. Rev. Lett. **85**, 2937 (2000).
4. A. I. Maĭmistov, Zh. Éksp. Teor. Fiz. **104**, 3620 (1993) [JETP **77**, 727 (1993)].
5. N. N. Bautin and E. A. Leontovich, *Methods and Techniques of Qualitative Investigation of Dynamic Systems on a Plane* (Nauka, Moscow, 1990) [in Russian].
6. N. N. Akhmediev, V. V. Afanasjev, and J. M. Soto-Crespo, Phys. Rev. E **53**, 1190 (1996).
7. N. Akhmediev, J. M. Soto-Crespo, and G. Town, Phys. Rev. E **63**, 056602 (2001).
8. A. A. Samarskiĭ and E. S. Nikolaev, *Numerical Methods for Grid Equations: Direct Methods* (Nauka, Moscow, 1978; Birkhäuser, Boston, 1991).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 2, 2004, pp. 443–448.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 2, 2004, pp. 502–508.

MISCELLANEOUS

Random Walks with Intermediate Anomalous-Diffusion Asymptotics

A. I. Saichev and S. G. Utkin
Nizhni Novgorod State University, Nizhni Novgorod, 603600 Russia
e-mail: sergei_utkin@mail.ru
Received March 12, 2004

Abstract—"Quasi-anomalous" random walks are considered that have linear-diffusion asymptotics at long times and obey anomalous-diffusion laws at intermediate times (which are also long as compared to microscopic time scales). A generalized fractional-exponential distribution with bounded moments is introduced. It is shown that random walks characterized by waiting-time distributions of this kind exhibit both normal and anomalous diffusion asymptotics. The correctness of analytical calculations is confirmed by numerical computations. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In recent years, numerous studies have been devoted to analysis of anomalous diffusion (e.g., see [1] for review and [2, 3]). Unlike linear diffusion, it is characterized by nonlinear growth of the mean square of a diffusion process with time, generally obeying the law

$$\langle X^2(t)\rangle \propto t^\gamma,$$

which reflects a violation of the law of large numbers and/or the central limit theorem. Anomalous diffusion phenomena are encountered in various physical problems, such as charge transfer in amorphous semiconductors [4–9], fractal geometry [10], quantum optics [11, 12], Richardson's law of turbulent diffusion [13–15], and chaotic dynamics of Hamiltonian systems [16]. The long-time and long-distance asymptotic behavior of such stochastic processes is adequately described by fractional partial differential equations. For this reason, one objective of this study is to construct random processes that provide adequate models of physical anomalous diffusion processes characterized by probability distributions that are exact solutions to the aforementioned fractional differential equations.

In addition to anomalous diffusion, of interest for various applications are "quasi-anomalous" random processes. They obey the law of large numbers and the central limit theorem at extremely long times and exhibit universal asymptotic behavior characteristic of anomalous diffusion at intermediate long times. In this study, we consider quasi-anomalous random walks of this kind exhibiting intermediate-time behavior of anomalous-diffusion type and obeying the linear law of diffusion as $t \longrightarrow \infty$.

2. MODEL OF RANDOM WALKS

Consider a typical random walk described by the stochastic equation

$$\frac{dX(t)}{dt} = \sum_k h_k \delta(t - t_k).$$

The process $X(t)$ is interpreted as the coordinate of a particle that jumps to distances h_k at random times t_k. Assume that the random variables h_k and $\tau_k = t_k - t_{k-1}$ are mutually independent and are characterized by distributions $w(x)$ and $f(\tau)$, respectively. It is obvious that

$$X(t) = \sum_{k=1}^{N(t)} h_k,$$

where $N(t)$ is the number of jumps executed by the time t. The function $N(t)$ is the inverse of the nth-jump time $T(n)$:

$$t = T(n) = \begin{cases} 0, & n = 0, \\ \sum_{k=1}^{N} \tau_k, & n \geq 1. \end{cases}$$

In other words, $N(t)$ and $T(t)$ obey an equivalence relation:

$$N(t) \geq n \longleftrightarrow T(n) < t. \qquad (1)$$

To derive an equation for the probability density $W(x, t)$ of the process $X(t)$, we begin with defining the

characteristic function

$$\Theta(u; t) = \langle \exp(iuX(t)) \rangle = \left\langle \exp\left(iu \sum_{k=1}^{N(t)} h_k \right) \right\rangle.$$

The last mean value is difficult to calculate, because the sum to be averaged contains a random number $N(t)$ of summands. To change from $n = N(t)$ to the better studied function $t = T(n)$, we make use of the following partition of unity:

$$1 = \sum_{n=0}^{\infty} \Pi_n(z) = \sum_{n=0}^{\infty} [\chi(z-n) - \chi(z-n-1)], \quad z > 0,$$

where $\chi(x)$ is the Heaviside unit step function. Accordingly, the desired characteristic function takes the form

$$\Theta(u; t) = \sum_{n=0}^{\infty} \left\langle \exp\left(iu \sum_{k=1}^{n} h_k \right) \Pi_n(N(t)) \right\rangle.$$

Using (1) to express $\Pi_n(N(t))$ in terms of $T(n)$, we obtain

$$\Theta(u; t) = \sum_{n=0}^{\infty} \left\langle \exp\left(iu \sum_{k=1}^{n} h_k \right) [\chi(t - T(n)) - \chi(t - T(n+1))] \right\rangle.$$

The sum of the geometric progression obtained as the Laplace transform of this expression is the Montroll–Weiss equation [17]:

$$\hat{\Theta}(u; s) = \frac{1 - \hat{f}(s)}{s[1 - \tilde{w}(u)\hat{f}(s)]}. \tag{2}$$

Here, $\hat{f}(s)$ is the Laplace transform of the waiting-time distribution and $\tilde{w}(u)$ is the jump-size characteristic function. This equation is equivalent to

$$\frac{1}{\hat{f}(s)}\hat{\Theta}(u; s) - \tilde{w}(u)\hat{\Theta}(u; s) = \frac{1 - \hat{f}(s)}{s\hat{f}(s)}. \tag{3}$$

In what follows, this equation is used to derive both the classical Kolmogorov–Feller kinetic equation and kinetic equations describing anomalous diffusion for different distributions $f(\tau)$ and $w(x)$. This equation is also used to analyze the transition from anomalous to linear diffusion for quasi-anomalous random walks.

3. FRACTIONAL-EXPONENTIAL DISTRIBUTION

Anomalous random walks are conveniently modeled by using a fractional-exponential distribution $\varphi_\beta(\tau)$ of intervals τ_k, which is represented in integral form below. Its Laplace transform is

$$\hat{\varphi}_\beta(s) = \frac{1}{1 + s^\beta}, \quad 0 < \beta < 1. \tag{4}$$

The corresponding distribution can be expressed in terms of the Mittag–Leffler function

$$E_\beta(z) = \frac{1}{2\pi i}\int_{\mathcal{H}} \frac{e^y y^{\beta-1}}{y^\beta - z} dy,$$

where the integral is calculated over the Hankel loop $\mathcal{H}$ encompassing the negative real axis [18, 19]. As a result, we have

$$\varphi_\beta(\tau) = -\frac{1}{\tau} D(-\tau^\beta)\chi(\tau), \quad D_\beta(z) = \beta z \frac{dE_\beta(z)}{dz}.$$

One can readily find the following integral representation of this distribution:

$$\varphi_\beta(\tau) = \frac{\sin(\pi\beta)}{\pi}\tau^{\beta-1}\int_0^\infty \frac{x^\beta e^{-x} dx}{x^{2\beta} + \tau^{2\beta} + 2x^\beta\tau^\beta\cos(\pi\beta)}, \quad 0 < \beta < 1,$$

which behaves as $\tau^{\beta-1}$ as $\tau \longrightarrow 0$ and $\tau^{-\beta-1}$ as $\tau \longrightarrow \infty$ [19]. The latter asymptotic of the fractional-exponential distribution implies that its mean value is infinitely large ($\langle\tau\rangle = \infty$). Furthermore, the conditions of the law of large numbers are violated for the function $T(N)$, and the probability density of $X(t)$ obeys a fractional partial differential equation. Indeed, substituting (4) into (3), we obtain the generalized Kolmogorov–Feller equation

$$\frac{\partial^\beta W(x; t)}{\partial t^\beta} + [W(x; t) - W(x; t) * w(x)] = \frac{t^{-\beta}}{\Gamma(1-\beta)}\chi(t)\delta(x),$$

which contains a fractional time derivative. Note also that the integral operator in x can be replaced by a differential one at long times, and the Kolmogorov–Feller equation reduces to an anomalous diffusion equation [17, 19]:

$$\frac{\partial^\beta W(x; t)}{\partial t^\beta} = \frac{\sigma^2}{2}\frac{\partial^2 W(x; t)}{\partial x^2} + \frac{\tau^{-\beta}}{\Gamma(1-\beta)}\chi(t)\delta(x),$$

where

$$\sigma^2 = \langle h^2 \rangle = \int_{-\infty}^{\infty} x^2 w(x) dx$$

is the jump-size variance.

4. GENERALIZED FRACTIONAL-EXPONENTIAL DISTRIBUTION

From a physical perspective, the infinitely large mean waiting time for the process $X(t)$ is a disadvantage of the fractional-exponential distribution, because the observed random intervals τ_k are characterized by a finite mean value. Moreover, an infinite mean implies that the law of large numbers is violated and anomalous diffusion must be observed. It is shown below that physical systems may exhibit anomalous diffusive regimes even in the case of a bounded mean value. To do this, we make use of the distribution $\varphi(\tau)$ whose Laplace transform is

$$\hat{\varphi}_{\beta,\delta}(s) = \frac{1+\delta^\beta}{1+(s+\delta)^\beta}. \quad (5)$$

It can readily be shown that the distribution with Laplace transform (5) is expressed in terms of a fractional-exponential distribution as

$$\varphi_{\beta,\delta}(\tau) = (1+\delta^\beta)e^{-\delta\tau}\varphi_\beta(\tau).$$

The distribution ($\varphi(\tau)$) has bounded moments and reduces to fractional-exponential distribution (4) as $\delta \longrightarrow 0$. Note that distributions of this kind have never been considered in studies of anomalous diffusion. It can only be recalled that, considering the case of divergent moments of displacement and waiting time, the authors of [2, 3] remarked that "if they are finite, then the effective transport equation asymptotically transforms into the classical diffusion equation (i.e., at macroscopic times $t \gg \langle t\rangle$ and distances $|x| \gg \sqrt{\langle x^2\rangle}$)."

Note that (5) is the Laplace transform

$$\hat{\varphi}(s) = \frac{w^\beta+\delta^\beta}{w^\beta+(s+\delta)^\beta}$$

reduced to dimensionless form, where both $1/w$ and $1/\delta$ have the dimensionality of time. Behavior of the system at times shorter than $1/w$ depends on the fine structure of the distributions $\varphi(\tau)$ and $w(x)$ and does not obey any universal diffusion law. On the other hand, if δ is so small that the interval between $1/w$ and $1/\delta$ is sufficiently long, then two variants of asymptotic behavior of a random process are possible.

In the case of $s \ll w$ considered in detail below, corollaries to general equation (3) can be analyzed by using Taylor series expansion in terms of s. Since δ is a small parameter, either $s \ll \delta \ll w$ or $\delta \ll s \ll w$. Both cases are considered on a macroscopic time scale much greater than the "microscopic" time $1/w$.

5. DIFFUSION EQUATIONS

To derive equations for the probability density of $X(t)$, we substitute the Laplace transform $\hat{\varphi}(s)$ given by (5) into Eq. (3) and consider the asymptotic behavior as $s \longrightarrow 0$, i.e., long-time asymptotics, in which case the Laplace transforms of distributions can be represented as Taylor series. We analyze in detail the case of $\delta \ll 1$, which is of primary importance for theory of anomalous diffusion. In this case, the random walk $X(t)$ exhibits properties characteristic of both anomalous and linear diffusion.

Two cases are considered here regarding the distribution $\varphi(\tau)$ and its Laplace transform (5): $s \ll \delta \ll 1$, which corresponds to $t \longrightarrow \infty$, and $\delta \ll s \ll 1$ (i.e., $1 \ll t \ll 1/\delta$), which corresponds to an intermediate regime. In the former case, $\hat{\varphi}(s) \sim 1 - Ds$, where $D = \beta\delta^{\beta-1}/(1+\delta^\beta)$, and (3) becomes

$$s\hat{\Theta}(u;s) + \frac{1}{D}[1-\tilde{w}(u)]\hat{\Theta}(u;s) = 1.$$

In the latter case, $\hat{\varphi}(s) \sim 1 - D's^\beta$, where $D' = (1+\delta^\beta)^{-1}$, and we have

$$s^\beta\hat{\Theta}(u;s) + \frac{1}{D'}[1-\hat{w}(u)]\hat{\Theta}(u;s) = s^{\beta-1}.$$

Applying the inverse Fourier and Laplace transforms to these equations, we obtain the Kolmogorov–Feller equation

$$\frac{\partial W(x;t)}{\partial t} + \frac{1}{D}[W(x;t) - W(x;t) * w(x)] = \delta(t)\delta(x),$$

$$t \longrightarrow \infty,$$

or the generalized Kolmogorov–Feller equation

$$\frac{\partial^\beta W(x;t)}{\partial t^\beta} + \frac{1}{D'}[W(x;t) - W(x;t) * w(x)] = \frac{t^{-\beta}}{\Gamma(1-\beta)}\chi(t)\delta(x),$$

$$1 \ll t \ll \frac{1}{\delta}.$$

To be specific, we substitute the jump-size distribution $w(x)$ whose Fourier transform is

$$\tilde{w}(u) \sim 1 - \frac{\sigma^2u^2}{2}, \quad u \longrightarrow 0$$

(as in the case of the Gaussian distribution) into these equations. Then, we obtain linear and anomalous diffusion equations describing different time-asymptotic

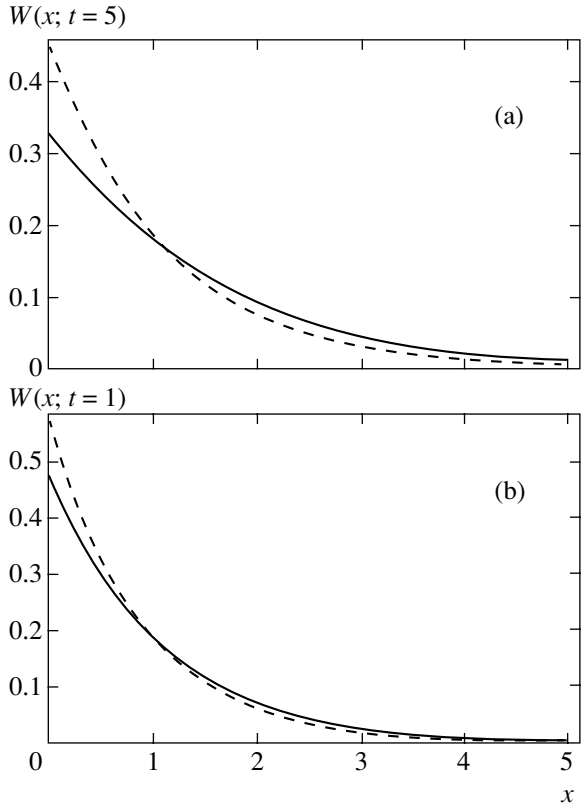


Fig. 1. Probability distributions at $t = 5$ (a) and 1 (b) ($\ll 1/\delta$) for $\delta = 0.01$, $\beta = 0.3$, and $\sigma = 1$. The dashed curve is the asymptotic solution expressed in terms of the Mittag–Leffler function; the solid curve is the exact solution.

regimes:

$$\frac{\partial W(x;\,t)}{\partial t} = \frac{\sigma^2}{2D}\frac{\partial^2 W(x;\,t)}{\partial x^2} + \delta(t)\delta(x), \quad t \longrightarrow \infty,$$

$$\frac{\partial^\beta W(x;\,t)}{\partial t^\beta} = \frac{\sigma^2}{2D'}\frac{\partial^2 W(x;\,t)}{\partial x^2} + \frac{t^{-\beta}}{\Gamma(1-\beta)}\chi(t)\delta(x),$$

$$1 \ll t \ll \frac{1}{\delta}.$$

The former equation has the well-known solution

$$W(x;\,t) = \frac{1}{\sqrt{2\pi\sigma^2 t/D}}\exp\left(-\frac{x^2}{2\sigma^2 t/D}\right).$$

The latter can be solved by using the model process of "fractional advection" described in [19]. The result is

$$W(x;\,t) = \frac{1}{\sqrt{2\sigma^2 t^\beta/D'}}Q_{\beta/2}\left(\frac{2|x|}{\sqrt{2\sigma^2 t^\beta/D'}}\right),$$

where the function $Q_\gamma(\tau)$ is found by applying the inverse Fourier transform to the Mittag–Leffler function $E_\gamma(iu)$.

6. COMPARISON OF EXACT AND ASYMPTOTIC SOLUTIONS TO THE MONTROLL–WEISS EQUATION

The asymptotic distributions obtained above for a quasi-anomalous random walk $X(t)$ correspond to different time intervals. Now, we should question the consistency of asymptotic solutions with the actual behavior of the process and determine the time intervals on which the analytical results obtained above are valid. This can be done by performing an accurate numerical analysis of the Montroll–Weiss equation (2).

Substituting Laplace transform (5) of the generalized fractional-exponential distribution into Eq. (2), we obtain the Laplace transform of the characteristic function in explicit form:

$$\hat\Theta(u;\,s) = \frac{1}{s}\frac{1}{\dfrac{1+\delta^\beta}{(s+\delta)^\beta - \delta^\beta}\dfrac{\sigma^2 u^2}{2}}. \tag{6}$$

To find an expression for the probability density $W(x, t)$ of the process $X(t)$, we apply the inverse Fourier transform to $\hat\Theta(u; s)$ to obtain

$$\hat W(x;\,s) = \frac{1}{\sqrt{2}\sigma s}\sqrt{\frac{(s+\delta)^\beta - \delta^\beta}{1+\delta^\beta}}$$
$$\times \exp\left(-\left|\frac{x\sqrt{2}}{\sigma}\sqrt{\frac{(s+\delta)^\beta - \delta^\beta}{1+\delta^\beta}}\right|\right);$$

we then apply the inverse Laplace transform in s. In the case under analysis, the Mellin integral can be reduced to an integral over the negative real axis. The result is

$$W(x;\,t) = \frac{1}{\pi\sigma\sqrt{2(1+\delta^\beta)}}$$
$$\times \mathrm{Im}\left\{e^{-i\pi}\int_0^\infty \frac{\sqrt{(ye^{i\pi}+\delta)^\beta - \delta^\beta}}{y}\right.$$
$$\left.\times \exp\left(-yt - \frac{x\sqrt{2}}{\sigma\sqrt{1+\delta^\beta}}\sqrt{(ye^{i\pi}+\delta)^\beta - \delta^\beta}\right)dy\right\}.$$

This exact formula is valid at any time. By analogy, we can use the integral representation of the Mittag–Leffler function to express $W(x, t)$ in an asymptotic integral form at "intermediate" times $1 \ll t \ll 1/\delta$:

$$W(x;\,t) = \frac{1}{\pi\sigma\sqrt{2(1+\delta^\beta)}}\,\mathrm{Im}\left\{\exp\left(\frac{i\pi\beta}{2}\right)\int_0^\infty \exp\left(-yt - \frac{x\sqrt{2}}{\sigma\sqrt{1+\delta^\beta}}y^{\frac{\beta}{2}}\exp\left(\frac{i\pi\beta}{2}\right)\right)\frac{dy}{y^{1-\beta/2}}\right\}.$$

These formulas were used to calculate the asymptotic and exact probability densities at long times ($t \longrightarrow \infty$, Fig. 1) and intermediate times ($1 \ll t \ll 1/\delta$, Fig. 2) for the same values of β, δ, and σ. Figure 1 demonstrates that the exact solution tends to the asymptotic distribution for anomalous diffusion with decreasing time. The long-time behavior exhibits an analogous trend (see Fig. 2): the exact solution approaches the asymptotic Gaussian distribution with increasing time.

It is also interesting to examine asymptotic behavior of the mean square $\langle X^2(t)\rangle$. First, we find its Laplace transform:

$$g(s) = \int \langle X^2(t)\rangle e^{st}dt = \left.\frac{\partial^2\hat{\Theta}(u;\,s)}{\partial u^2}\right|_{u=0}. \qquad (7)$$

Substituting the solution to the Montroll–Weiss equation (2) with $\tilde{w}(u) \sim 1 - \sigma^2u^2/2$ and either $\hat{f}(s) \sim 1 - Ds$ (if $s \ll \delta \ll 1$) or $\hat{f}(s) \sim 1 - D's^\beta$ (if $\delta \ll s \ll 1$) into this expression, we obtain

$$g(s) \sim \frac{\sigma^2}{Ds^2}, \quad s \ll \delta \ll 1,$$

$$g(s) \sim \frac{\sigma^2}{D's^{1+\beta}}, \quad \delta \ll s \ll 1.$$

Accordingly, the behavior of the mean square at long and intermediate times is as follows:

$$\langle X^2(t)\rangle \sim \frac{\sigma^2}{D}t, \quad t \longrightarrow \infty,$$

$$\langle X^2(t)\rangle \sim \frac{\sigma^2}{D'\Gamma(1+\beta)}t^\beta, \quad 1 \ll t \ll \frac{1}{\delta}.$$

To examine the exact behavior of $\langle X^2(t)\rangle$, we substitute $\hat{\Theta}(u;\,s)$ given by (6) into (7):

$$g(s) = \frac{\sigma^2(1+\delta^\beta)}{s[(s+\delta)^\beta - \delta^\beta]}.$$

An exact expression for the mean square can be found for $\beta = 1/2$:

$$\langle X^2(t)\rangle = \frac{1+\sqrt{\delta}}{\sqrt{\delta}} \times\left[\sqrt{\frac{\delta t}{\pi}}e^{-\delta t} + \delta t + \left(\delta t + \frac{1}{2}\right)\mathrm{erf}(\sqrt{\delta t})\right].$$

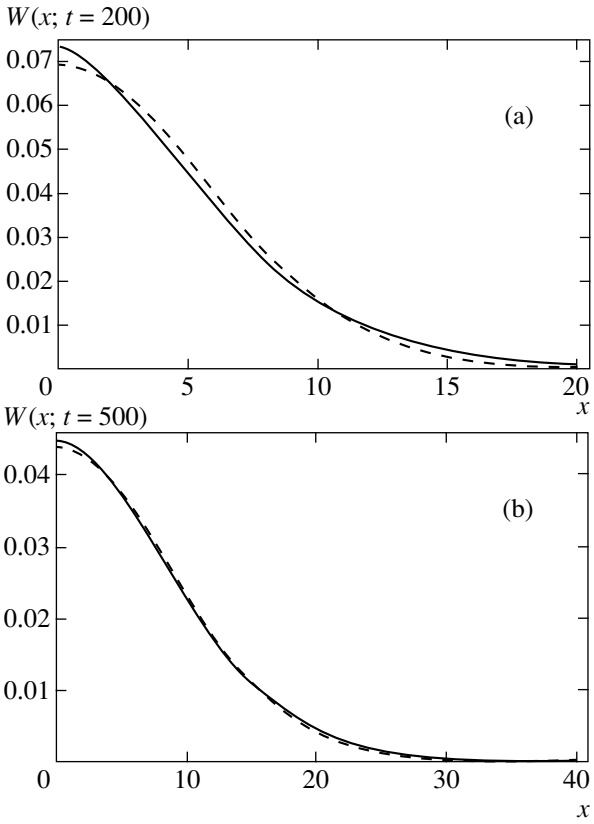


Fig. 2. Probability distributions at $t = 200$ (a) and 500 (b) ($\geqslant 1/\delta$) for $\delta = 0.01$, $\beta = 0.3$, and $\sigma = 1$. The dashed curve is the asymptotic solution expressed in terms of the Mittag–Leffler function; the solid curve is the exact numerical solution.

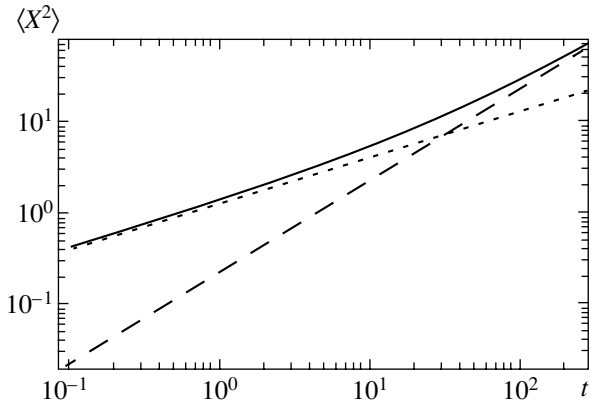


Fig. 3. Asymptotic behavior of the mean square of the quasi-anomalous diffusion process at long times (dashed curve) and intermediate times (dotted curve) and the exact $\langle X^2(t)\rangle$ curve for $\beta = 1/2$ (solid curve).

Figure 3 compares this function with the corresponding intermediate- and long-time asymptotics (corresponding to times on the order of ($\sim\sqrt{t}$) and ($\sim t$), respectively). It is clear that the graph of the exact solution agrees with the respective asymptotic curves at long and intermediate times.

Thus, we have shown that anomalous diffusion can take place in the case of a finite mean waiting time $\langle\tau_k\rangle$. However, anomalous diffusion of this kind corresponds to an intermediate asymptotic regime followed by normal linear diffusion.

ACKNOWLEDGMENTS

This work was supported by the NSh-838.2003.2 under the Presidential Program of State Support of Leading Scientific Schools; by the Russian Foundation for Basic Research, project no. 03-02-16680; and by the Ministry of Education of the Russian Federation, project no. E02-3.5-232.

REFERENCES

1. R. Metzler and J. Klafter, Phys. Rep. **339**, 1 (2000).
2. K. V. Chukbar, Zh. Éksp. Teor. Fiz. **108**, 1875 (1995) [JETP **81**, 1025 (1995)].
3. V. Yu. Zaburdaev and K. V. Chukbar, Zh. Éksp. Teor. Fiz. **121**, 299 (2002) [JETP **94**, 252 (2002)].
4. V. V. Uchaĭkin, Zh. Tekh. Fiz. **68** (1), 138 (1998) [Tech. Phys. **43**, 124 (1998)].
5. V. V. Uchaĭkin, Teor. Mat. Fiz. **115**, 154 (1998).
6. V. V. Uchaĭkin and V. V. Saenko, Zh. Tekh. Fiz. **71** (2), 8 (2001) [Tech. Phys. **46**, 139 (2001)].
7. P. W. M. Blom and M. C. J. M. Vissenberg, Phys. Rev. Lett. **80**, 3819 (1998).
8. Q. Gu, E. A. Schiff, and S. Grebner, Phys. Rev. Lett. **76**, 3196 (1996).
9. M. Porto, A. Bunde, and S. Havlin, Phys. Rev. E **56**, 1667 (1997).
10. B. Rinn, W. Dieterich, and P. Maass, Philos. Mag. B **77**, 1283 (1998).
11. S. Schaufler, W. P. Schleich, and V. P. Yakovlev, Phys. Rev. Lett. **83**, 3162 (1999).
12. S. Schaufler, W. P. Schleich, and V. P. Yakovlev, Europhys. Lett. **39**, 383 (1997).
13. M. F. Shlesinger, B. J. West, and J. Klafter, Phys. Rev. Lett. **58**, 1100 (1987).
14. I. M. Sokolov, A. Blumen, and J. Klafter, Europhys. Lett. **47**, 152 (1999).
15. G. M. Zaslavsky and S. Benkadda, *Chaos, Kinetics and Nonlinear Dynamics in Fluids and Plasmas* (Springer, Berlin, 1998).
16. G. M. Zaslavsky, M. Edelman, and B. Niyazov, Chaos **7**, 159 (1997).
17. A. I. Saichev and G. M. Zaslavsky, Chaos **7**, 753 (1997).
18. A. I. Saichev and W. A. Woyczyński, *Distributions in the Physical and Engineering Sciences* (Birkhäuser, Boston, 1997), Vol. 1, p. 336.
19. A. I. Saichev and S. G. Utkin, Aktual. Probl. Stat. Radiofiz. **1**, 5 (2002).

Translated by A. Betev