Journal of Experimental and Theoretical Physics, Vol. 100, No. 6, 2005, pp. 1043–1049.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 6, 2005, pp. 1187–1194.
Original Russian Text Copyright © 2005 by Matveev, Gusarevich, Pashev.

ATOMS, MOLECULES, OPTICS

Inelastic Processes in the Interaction of an Atom with an Ultrashort Electromagnetic Pulse

V. I. Matveev*, E. S. Gusarevich, and I. N. Pashev
Lomonosov Pomorskiĭ State University, Arkhangel'sk State Technical University, Arkhangel'sk, 163006 Russia
**e-mail: matveev.victor@pomorsu.ru*

Received May 31, 2004

Abstract—Electron transitions occurring during the interaction of a heavy relativistic atom with a spatially inhomogeneous ultrashort electromagnetic pulse are considered by solving the Dirac equation. The corresponding transition probabilities are expressed in terms of known inelastic atomic form factors, which are widely used in the theory of relativistic collisions between charged particles and atoms. By way of example, the inelastic processes accompanying the interaction of ultrashort pulses with hydrogen-like atoms are considered. The probabilities of ionization and production of a bound–free electron–positron pair on a bare nucleus, which are accompanied by the formation of a hydrogen-like atom in the final state and a positron in the continuum, are calculated. The developed technique makes it possible to take into account exactly not only the spatial inhomogeneity of an ultrashort electromagnetic pulse, but also the magnetic interaction. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The interaction of atoms with ultrashort electromagnetic pulses with a duration shorter than the characteristic atomic periods of time has become an object of investigation only recently. A new trend, viz., the physics of attosecond pulses (1 attos = 10^{-18} s), has been developed. The possibility of detection, generation, and application of attosecond pulses was discussed by many authors engaged in experimental and theoretical studies. The state of the art by the beginning of 2004 and the corresponding references are given in reviews [1–3] (see also several later publications [4–20]). The increased interest in the physics of ultrashort pulses is associated not only with modern tendencies in designing more powerful lasers and generation of ultrashort pulses [21], but also with the advances made in heavy-ion accelerator technique since the fields produced by relativistic and ultrarelativistic charged particles are close in properties to the field of a light wave. For example, in experiments [22] (see also [23–26]), double and single ionization of a helium atom by an impact of a uranium U^{92+} ion with an energy of 1 GeV/nucleon was studied and a ultrastrong pulse ($I > 10^{19}$ W/cm^2) with a duration on the order of 10^{-18} s was simulated. It is extremely difficult to obtain such parameters of an electromagnetic pulse by other available methods. For example, the observation of pulses with a duration of a few femtosecond was reported in [21], representing almost thirty years (up to 2000) of evolution in the physics of ultrashort laser pulses and technological achievements in the field of generating such pulses. Thus, collision experiments in fact offer the only possibility of simulating ultrashort pulses with a duration comparable to or smaller than the characteristic atomic time $\tau_a \sim 10^{-17}$ s. Collision experiments also provide the subsequent opportunity for direct observation of the interaction between atoms and an ultrashort electromagnetic pulse. In a comparatively recent experiment [27], multiphoton production of pairs by an ultrarelativistic electron moving with a relativistic factor of $\gamma \sim 10^5$ through an ultrastrong laser field was observed; in this case, in the rest system of the electron, the laser field frequency and strength increased approximately by a factor of γ. In recent theoretical works [28, 29], the processes of multiphoton pair production during collisions of bare ultrarelativistic nuclei with high-intensity laser radiation were considered and the possibility of conducting the corresponding experiments on modern accelerators was also noted. Thus, during the collision (interaction) of a target atom moving with a relativistic energy (or a partly stripped atom, viz., a structural ion with a certain number of electrons in its shells) with an ultrashort electromagnetic pulse of duration τ, the corresponding collision time τ_c in the rest system of the atom (ion) decreases by a factor of γ; i.e., $\tau_c \sim \tau/\gamma$. Let us consider the possibility of observing in such experiments the inelastic processes accompanying the interaction between atoms and ultrashort electromagnetic pulses with values of relativistic factor $\gamma \sim 10^4$ attainable in modern heavy-particle accelerators [24] (these values correspond to the effective decrease in the pulse duration by four orders of magnitude). We will first obtain estimates for relativistic problems, in which the characteristic energy difference $\Delta E \sim mc^2$ (m is the electron mass and c is the velocity of light). The corresponding characteristic frequency is $\omega_a = mc^2/\hbar$; consequently, the characteristic times of a stationary target atom are

$$\tau_a = \frac{2\pi}{\omega_a} \approx 8.1 \times 10^{-21} \sim 10^{-20} \text{ s},$$

while the collision time in the rest system of the atom for femtosecond pulses of duration $\tau \sim 10^{-15}$ s attainable at present is $\tau_c \sim \tau/\gamma \sim 10^{-19}$ s. Thus, direct observation of the relativistic effects considered here requires that the pulse duration be reduced by an order of magnitude (i.e., to approximately 100 attos), which is in line with contemporary tendencies [1–3, 21].

In theoretical analysis of the effects accompanying the interaction of atoms with ultrashort electromagnetic pulses, a natural foundation for solving problems can be the sudden approximation, which is closely related [30] to the eikonal approximation and which was previously used only for solving nonrelativistic problems [31–35], in which the perturbation is not small enough for using perturbation theory, but the time of action of a perturbation is much shorter than the characteristic periods of time for an unperturbed system. This makes it possible to solve the problem without setting a limit on the perturbation intensity. The effects of interaction of atoms with ultrashort electromagnetic pulses can be attributed to such cases. Here, we apply the term ultrashort pulses to pulses with duration smaller than the characteristic times for the target atom, which can be in the ground state or in an excited state (including a highly excited Rydberg state) prior to the interaction. Such pulses may be of various origin [1–3, 36–39], but can also be the fields of heavy ions moving with a relativistic or ultrarelativistic velocity [22–26]. In the latter case, perturbation theory is inapplicable for the fields of ions with large charges [40] even for infinitely large energies of the ions. A nonrelativistic nonperturbative theory developed in [41] describes the electron transitions and radiation emitted by an atom during its interaction with a spatially inhomogeneous (over the target size) ultrashort electromagnetic pulse.

In this study, on the basis of the sudden approximation, we obtain a solution to the Dirac equation, which describes the behavior of a hydrogen-like atom during its interaction with a spatially inhomogeneous ultrashort electromagnetic pulse. The corresponding transition probabilities are expressed in terms of the known inelastic atomic form factors, which are widely used in the theory of relativistic collisions between charged particles and atoms. By way of example, we consider the inelastic processes accompanying the interaction between ultrashort pulses with hydrogen-like atoms; the probabilities of ionization and production of a bound–free electron–positron pair at a bare nucleus accompanied by the formation of a hydrogen-like atom in the final state and a positron in the continuum are calculated. The developed technique makes it possible to exactly take into account the spatial inhomogeneity (over the target size) of the ultrashort electromagnetic pulse as in the nonrelativistic theory [41]; however, in contrast to the nonrelativistic theory, this technique exactly takes into account the magnetic interaction also.

2. TRANSITION AMPLITUDE IN THE SUDDEN APPROXIMATION

In the terminology used in [31], the perturbation corresponding to the field of an ultrashort pulse is a shock of the scattering type. To illustrate the sudden approximation, it is obviously expedient to consider the formal solution of the Schrödinger equation (here and below, we use atomic units)

$$i\dot{\Psi} = (H_0 + U(t))\Psi, \tag{1}$$

where sudden perturbation $U(t)$ acts during a time much shorter than the characteristic time periods of an unperturbed system described by Hamiltonian H_0. In this case, in solving Eq. (1), we can disregard (during the time of action of perturbation $U(t)$) the evolution of the wavefunction under the action of intrinsic Hamiltonian H_0 and solve the equation

$$i\dot{\Psi} = U(t)\Psi.$$

It hence follows that

$$\Psi(t) = \exp\left\{-i\int_{t_0}^{t} U(t)dt\right\}\Psi(t_0). \tag{2}$$

Consequently, the amplitude of transition of a nonrelativistic atom from state $|i\rangle$ to state $|f\rangle$ as a result of sudden perturbation $U(t)$ has the form [31]

$$a_{if} = \langle f|\exp\left\{-i\int_{-\infty}^{+\infty} U(t)dt\right\}|i\rangle. \tag{3}$$

It can easily be seen that the same result can be obtained if we solve exactly Eq. (1) with a delta-shaped potential $\tilde{U}(t)$ connected with potential $U(t)$ via the relation

$$\tilde{U}(t) = U_0\delta(t), \quad U_0 = \int_{-\infty}^{+\infty} U(t)dt. \tag{4}$$

Precisely this circumstance will be used below for solving the Dirac equation in the sudden approximation.

The behavior of the electron in a hydrogen-like atom (with a nuclear charge Z_a on which no limitations are imposed except the applicability conditions [42] for the Dirac equation) in an external field

$$A^\mu = (\varphi, \mathbf{A})$$

will be described by the Dirac equation (the electron charge $e = -1$ at. unit)

$$i\dot{\Psi} = \left\{c\hat{\boldsymbol{\alpha}}\left(\hat{\mathbf{p}} + \frac{1}{c}\mathbf{A}\right) - \frac{Z_a}{r} - \varphi + \hat{\beta}c^2\right\}\Psi, \qquad (5)$$

where the terms

$$c\hat{\boldsymbol{\alpha}}\hat{\mathbf{p}} + \hat{\beta}c^2 - Z_a/r$$

are equal to the Hamiltonian H_0 of a single atom and the interaction of the atomic electron with the external field is described by the potential

$$U(t) = U(\mathbf{r}, t) = \hat{\boldsymbol{\alpha}}\mathbf{A} - \varphi,$$

where $\hat{\mathbf{p}}$ is the momentum operator, $\hat{\boldsymbol{\alpha}}$ and $\hat{\beta}$ are the Dirac matrices, and $\mathbf{r}$ are the coordinates of the atomic electron. We first choose the calibration of the electromagnetic wave potentials (vector potential $\mathbf{A}$ and scalar potential φ) so that the scalar potential is zero. We assume that the vector potential of the field of the wave is a function of coordinate $\mathbf{r}$ and time t,

$$\mathbf{A} = \mathbf{A}(\mathbf{r}, t) = \mathbf{A}(\eta),$$

where the phase of the wave is given by

$$\eta = \omega_0 t - \mathbf{k}_0 \cdot \mathbf{r};$$

here, wavevector $\mathbf{k}_0$ is such that

$$|\mathbf{k}_0| = \omega_0/c,$$

and ω_0 is the circular frequency. We carry out the gauge transformation [41]

$$\mathbf{A}' = \mathbf{A} + \nabla f, \quad \varphi' = \varphi - \frac{1}{c}\frac{\partial f}{\partial t},$$

where

$$f = -\mathbf{A} \cdot \mathbf{r}.$$

This gives

$$\mathbf{A}' = \mathbf{k}_0\left(\mathbf{r}\frac{d\mathbf{A}}{d\eta}\right), \quad \varphi' = -(\mathbf{E} \cdot \mathbf{r}),$$

where

$$\mathbf{E} = \mathbf{E}(\mathbf{r}, t) = -|\mathbf{k}_0|\frac{d\mathbf{A}}{d\eta}.$$

Consequently, in the new calibration, the vector and scalar potentials are connected through the relation

$$\mathbf{A}' = (\mathbf{k}_0/|\mathbf{k}_0|)\varphi'.$$

We assume that the z axis is directed along vector $\mathbf{k}_0$. In this case, the interaction of an atomic electron with the external field in Eq. (5) is given by

$$\begin{aligned} U(t) &= \hat{\boldsymbol{\alpha}}\mathbf{A}' - \varphi' \\ &= -\left(1 - \frac{\hat{\boldsymbol{\alpha}}\mathbf{k}_0}{|\mathbf{k}_0|}\right)\varphi' = -(1 - \hat{\alpha}_z)\varphi'. \end{aligned} \qquad (6)$$

We will operate with the new calibration and omit primes on the potentials. To solve the Dirac equation (5) in the sudden approximation, we write it in the form

$$i\dot{\Psi} = (H_0 + U(t))\Psi$$

and use the substitution (4) introduced at the beginning of this section. For this purpose, we introduce

$$\tilde{\varphi} = \varphi_0\delta(ct - z), \quad \varphi_0 = c\int_{-\infty}^{+\infty}\varphi dt, \qquad (7)$$

where $\varphi_0 = \varphi_0(\mathbf{r})$ (i.e., a function of only the coordinates $\mathbf{r}$ of the point of observation). Further, in accordance with formula (4), we replace $U(t)$ from formula (6) by

$$\tilde{U}(t) = -(1 - \hat{\alpha}_z)\tilde{\varphi}$$

or by

$$\tilde{U}(t) = U_0\delta(ct - z), \qquad (8)$$

where

$$U_0 = c\int_{-\infty}^{+\infty}U(t)dt = -(1 - \hat{\alpha}_z)\varphi_0. \qquad (9)$$

As a result, Eq. (5) assumes the form

$$i\dot{\Psi} = \left\{c\hat{\boldsymbol{\alpha}}\hat{\mathbf{p}} - \frac{Z_a}{r} + \hat{\beta}c^2 - (1 - \hat{\alpha}_z)\tilde{\varphi}\right\}\Psi. \qquad (10)$$

To obtain the exact solution to the Dirac equation with such a potential, we expand $\Psi = \Psi(\mathbf{r}, t)$ in the eigenfunctions $\phi_k(\mathbf{r})$ (with energies E_k) of the unperturbed atomic Hamiltonian

$$H_0 = c\hat{\boldsymbol{\alpha}}\hat{\mathbf{p}} + \hat{\beta}c^2 - Z_a/r.$$

This gives

$$\Psi(\mathbf{r}, t) = \sum_k a_k(t)\phi_k(\mathbf{r})\exp(-iE_k t).$$

Substituting this expansion into the left-hand side of the equation

$$i\dot{\Psi} = (H_0 + \tilde{U}(t))\Psi$$

and integrating, after premultiplying it by a state ϕ_f and taking into account the orthogonality of states ϕ_k, we obtain

$$\frac{da_f(t)}{dt} = -i\exp(iE_f t)\langle\phi_f|\tilde{U}(t)|\Psi(\mathbf{r}, t)\rangle. \tag{11}$$

Let us suppose that the atom was in state ϕ_j before the collision; in this case, we have

$$\Psi(\mathbf{r}, t = -\infty) = \exp(-iE_j t)\phi_j(r), \tag{12}$$

$$a_f(t = -\infty) = \delta_{fj}, \tag{13}$$

where δ_{fj} is the Kronecker delta. Since

$$\tilde{U}(t) = U_0\delta(ct - z),$$

it is sufficient for solving Eq. (11) to know the values of $\Psi(\mathbf{r}, t)$ only for $ct = z$; these values can be determined from Eq. (10) as follows. We pass to the light-cone variables

$$\begin{aligned} z^- &= (ct - z), \\ z^+ &= (ct + z). \end{aligned} \tag{14}$$

Retaining only the derivatives with respect to z^- in a small neighborhood of $z^- = 0$ and the singular potential $-(1 - \hat{\alpha}_z)\tilde{\varphi}$, we obtain the equation

$$ic(1 - \hat{\alpha}_z)\frac{\partial\Psi}{\partial z^-} = -(1 - \hat{\alpha}_z)\tilde{\varphi}\Psi. \tag{15}$$

Since $\tilde{\varphi} = \varphi_0\delta(z^-)$, taking into account the relations

$$\begin{aligned} \frac{d}{dx}\theta(x) &= \delta(x), \\ \frac{d}{dx}\exp(\theta(x)) &= \delta(x)\exp(\theta(x)), \end{aligned} \tag{16}$$

where

$$\theta(x) = \begin{cases} 0, & x < 0, \\ 1, & x > 0, \end{cases}$$

we obtain the solution to Eq. (15):

$$\begin{aligned} (1 - \hat{\alpha}_z)\Psi(z^- + 0) &= (1 - \hat{\alpha}_z) \\ \times \exp\left[i\frac{\varphi_0}{c}\theta(z^-)\right]&\Psi(z^- - 0). \end{aligned} \tag{17}$$

Returning to time t and using conditions (12) and (13), we obtain a solution which is valid for $t < z/c$ and in the vicinity of $t = z/c$ (i.e., $t = z/c + \varepsilon$, $\varepsilon > 0$ and is small):

$$\begin{aligned} (1 - \hat{\alpha}_z)\Psi(\mathbf{r}, t) &= (1 - \hat{\alpha}_z) \\ \times \exp\left[i\theta(ct - z)\frac{\varphi_0}{c}\right]&\exp[-iE_j t]\phi_j(\mathbf{r}). \end{aligned} \tag{18}$$

Substituting Eq. (18) into the right-hand side of Eq. (11) and integrating with respect to t taking into account initial conditions (13), we obtain

$$\begin{aligned} a_{fj} &= a_f(t = +\infty) \\ &= \delta_{fj} + i\int_{-\infty}^{+\infty} dt\exp(i(E_f - E_j)t) \\ &\times\langle\phi_f|\varphi_0\delta(z - ct)(1 - \hat{\alpha}_z)\exp\left[i\theta(ct - z)\frac{\varphi_0}{c}\right]|\phi_j\rangle. \end{aligned} \tag{19}$$

Using relations (16), we obtain

$$\begin{aligned} a_{fj} &= \delta_{fj} + \langle\phi_f|(1 - \hat{\alpha}_z)\exp\left(i\frac{(E_f - E_j)z}{c}\right) \\ &\times\left[\exp\left(i\frac{\varphi_0}{c}\right) - 1\right]|\phi_j\rangle. \end{aligned} \tag{20}$$

This is the required solution to the Dirac equation with potential

$$\tilde{U}(t) = U_0\delta(ct - z),$$

corresponding to the inclusion of potential (6) in the sudden approximation. The obtained expression can be written in a more convenient form by using the easily verifiable relation

$$\langle\phi_f|(1 - \hat{\alpha}_z)\exp\left(i\frac{(E_f - E_j)z}{c}\right)|\phi_j\rangle = \delta_{fj}. \tag{21}$$

This gives

$$\begin{aligned} a_{fj} &= \langle\phi_f|(1 - \hat{\alpha}_z)\exp\left(i\frac{(E_f - E_j)z}{c}\right) \\ &\times\exp\left(i\frac{\varphi_0}{c}\right)|\phi_j\rangle. \end{aligned} \tag{22}$$

To describe transitions in the case of the interaction of a complex multielectron atom with an ultrashort electromagnetic pulse, we proceed as follows. We assume that the states of atomic electrons are described as the products of one-electron wavefunctions and denote the energy of the electron with number a (where $a = 1, 2, \ldots, N$, N being the number of atomic electrons) by $E^{(a)}$ and the electron coordinates by $\mathbf{r}_a$. Then the natural generalization of amplitude (22) for a transition of the complex N-electron atom from the initial state $\phi_j = \phi_j(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$ with energy

$$E_j = \sum_{a=1}^{N} E_j^{(a)}$$

to the final state $\phi_f = \phi_f(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$ with energy

$$E_f = \sum_{a=1}^{N} E_f^{(a)}$$

has the form

$$a_{fj} = \langle\phi_f|\prod_{a=1}^{N}(1-\hat{a}_z^{(a)})\exp\left[\frac{i}{c}\sum_{a=1}^{N}(E_f^{(a)}-E_j^{(a)})z_a\right] \times \exp\left(\frac{i}{c}\sum_{a=1}^{N}\varphi_0(\mathbf{r}_a)\right)|\phi_j\rangle, \tag{23}$$

where matrix $\hat{\alpha}_z^{(a)}$ acts only on bispinor indices belonging to the atomic electron with number a.

3. TRANSITION PROBABILITIES

Let us consider the interaction of an atomic electron with a Gaussian electromagnetic pulse (with an effective duration on the order of λ^{-1}),

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0\exp\left(-\lambda^2\left(t-\frac{\mathbf{k}_0\cdot\mathbf{r}}{\omega_0}\right)^2\right) \times \cos(\omega_0 t - \mathbf{k}_0\cdot\mathbf{r}), \tag{24}$$

$$\varphi = -\mathbf{r}\cdot\mathbf{E}(\mathbf{r}, t), \quad \varphi_0 = c\mathbf{q}\cdot\mathbf{r},$$

where

$$\mathbf{q} = -\int_{-\infty}^{+\infty} dt\mathbf{E}(\mathbf{r}, t) = -\mathbf{E}_0\frac{\sqrt{\pi}}{\lambda}\exp\left(-\frac{\omega_0^2}{4\lambda^2}\right). \tag{25}$$

Recollecting that

$$\hat{\alpha}_z = \hat{\boldsymbol{\alpha}}\mathbf{k}_0/k_0$$

and introducing the vector

$$\mathbf{Q} = (Q_x, Q_y, Q_z) = (q, 0, \Omega_{fj}/c),$$

where

$$\Omega_{fj} = E_f - E_j,$$

we write a_{fj} from relation (22) in the form

$$a_{fj} = \langle\phi_f|\left(1-\frac{\hat{\boldsymbol{\alpha}}\mathbf{k}_0}{k_0}\right)\exp(i\mathbf{Q}\cdot\mathbf{r})|\phi_j\rangle. \tag{26}$$

Choosing now the z axis in the direction of vector $\mathbf{Q}$, we can write a_{fj} in the form

$$a_{fj} = \langle\phi_f|(1-\hat{\alpha}_z\cos\theta)\exp(iQz)|\phi_j\rangle - \langle\phi_f|\hat{\alpha}_x\sin\theta\exp(iQz)|\phi_j\rangle, \tag{27}$$

where

$$Q = \sqrt{q^2+\Omega_{fj}^2/c^2}, \quad \cos\theta = \Omega_{fj}\Big/\sqrt{c^2q^2+\Omega_{fj}^2},$$

$$\sin\theta = q\Big/\sqrt{q^2+\Omega_{fj}^2/c^2}.$$

Thus, let us now suppose that the atom was in state ϕ_j with energy E_j prior to the interaction (i.e., at $t = -\infty$) with the field of an ultrashort pulse; then the probability of finding the atom in state ϕ_f with energy E_f after the interaction (i.e., for $t = +\infty$) is given by

$$|a_{fj}|^2 = \frac{q^2}{Q^2}\left(\frac{q^2}{Q^2}|F^{fj}|^2 + |G_x^{fj}|^2\right). \tag{28}$$

Following [43–45], we have introduced the inelastic atomic form factors

$$F^{fj} = \langle\phi_f|\exp(iQz)|\phi_j\rangle = \frac{Qc}{\Omega_{fj}}\langle\phi_f|\hat{\alpha}_z\exp(iQz)|\phi_j\rangle,$$

$$G_x^{fj} = \langle\phi_f|\hat{\alpha}_x\exp(iQz)|\phi_j\rangle,$$

which are widely used in the theory of relativistic collisions of charged particles with atoms. Consequently, the above formulas make it possible [43–45] to determine the probabilities of excitation and ionization of a hydrogen-like atom interacting with an ultrashort electromagnetic pulse. These formulas can also be used for calculating the probability of production of an electron–positron pair during the interaction of a bare ion with an ultrashort electromagnetic pulse if we interpret this process as a transition of an electron from the states of the negative continuum (Dirac sea) to the states with a positive total energy of the hydrogen-like atom In all cases, we can use either the form factors (see, for example, [40, 45]) calculated using the so-called Coulomb–Dirac hydrogen-like wavefunctions for electrons and positrons, which leads to a complex numerical computation, or analytic expressions [40, 43–47] for the form factors determined with the Darwin quasi-relativistic wavefunctions and the Sommerfeld–Maue wavefunctions [40, 42, 44, 45, 48]. Strictly speaking, the quasi-relativistic functions are valid under the condition $Z_a \ll c$; if this inequality is violated, the results satisfactorily illustrate the behavior of the form factors qualitatively [40, 45].

The above formulas make it possible to calculate both the probabilities of inelastic processes having nonrelativistic analogs (excitation or ionization of an atom by an ultrashort electromagnetic pulse) and essentially relativistic effects (production of the electron–positron pairs). As in the nonrelativistic theory [41], the relativistic approach developed by us exactly takes into account the spatial inhomogeneity (over the target size) of the ultrashort electromagnetic pulse; however, unlike the nonrelativistic approach [41] (see also [5, 49]), it exactly takes into account the magnetic interaction also. Figures 1 and 2 show the ionization probabilities

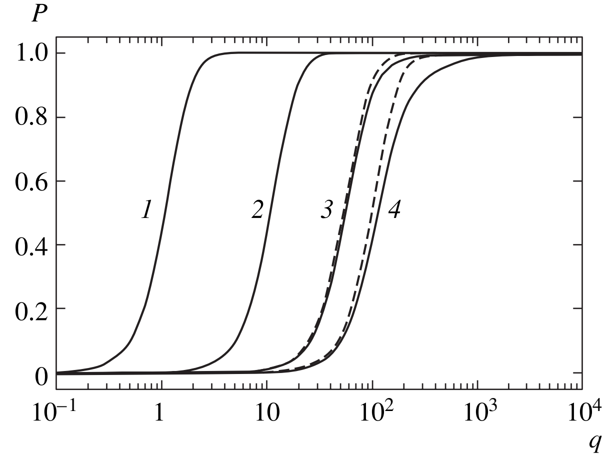


Fig. 1. Dependences of the ionization probability P (formation of a K vacancy) of hydrogen-like atoms for several values of nuclear charge Z_a on the transferred momentum (25) $q = |\mathbf{q}|$ (atomic units). The results of calculations for each value of Z_a are represented by two curves; the solid curve corresponds to relativistic calculations based on formula (28), while the dashed curve corresponds to the nonrelativistic calculation [41] (formula (28) for $c \longrightarrow \infty$) for $Z_a = 1$ (*1*), 10 (*2*), 50 (*3*), and 92 (*4*).

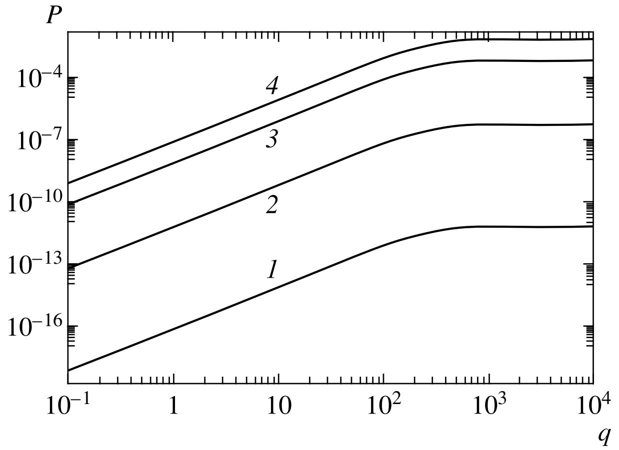


Fig. 2. Dependences of the probability P (28) of production of a free–bound electron–positron pair (the electron in the 1*s* state and the positron in the state of continuum of a hydrogen-like atom with effective atomic charge Z_a) on the transferred momentum (25) $q = |\mathbf{q}|$ (atomic units) for several values of $Z_a = 1$ (*1*), 10 (*2*), 50 (*3*), and 92 (*4*).

(the probabilities of formation of K vacancies) of hydrogen-like atoms and the probabilities of production of a free–bound electron–positron pair for several values of the nuclear charge Z_a. In our calculations, we used, following [40, 45], the quasi-relativistic Darwin wavefunctions for bound states and the Sommerfeld–Maue functions for the states of the continuum as the wavefunctions of the initial and final states; this makes it possible to calculate the form factors analytically.

To obtain the cross section of the transition of an atom from state ϕ_j with energy E_j to state ϕ_f with energy E_f, we must obviously multiply, according to [50], the probability $|a_{fj}|^2$ of the corresponding transition from formula (28) by the energy difference

$$\Omega_{fj} = E_f - E_j$$

and divide the result by the energy flux I equal to the integral of the absolute value of the Poynting vector

$$S(t) = c(4\pi)^{-1}\mathbf{E}^2$$

with respect to time, where $\mathbf{E}$ is expressed by formula (24). This gives

$$I = \int_{-\infty}^{+\infty} dt S(t) = \frac{c}{4\pi}\mathbf{E}_0^2\frac{\sqrt{\pi}}{2\sqrt{2}\lambda} \times\left\{\exp\left(-\frac{\omega_0^2}{2\lambda^2}\right)+1\right\}. \tag{29}$$

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project no. 04-02-16177) and INTAS (grant no. INTAS-GSI 03-54-4294).

REFERENCES

1. P. Pierre Agostini and L. F. Di Mauro, Rep. Prog. Phys. **67**, 813 (2004).
2. G. A. J. Reider, J. Phys. D: Appl. Phys. **37**, R37 (2004).
3. J. H. Posthumus, Rep. Prog. Phys. **67**, 623 (2004).
4. M. Boca, H. G. Muller, and M. Gavrila, J. Phys. B **37**, 147 (2004).
5. A. V. Lugovskoy and I. Bray, J. Phys. B **37**, 3427 (2004).
6. Th. Mercouris, Y. Komninos, and C. A. Nicolaides, Phys. Rev. A **69**, 032502 (2004).
7. A. Ben Haj Yedder, C. Le Bris, O. Atabek, *et al.*, Phys. Rev. A **69**, 041802 (2004).
8. K. T. Kim, C. M. Kim, M.-G. Baik, *et al.*, Phys. Rev. A **69**, 051805 (2004).
9. S. Kazamias and Ph. Balcou, Phys. Rev. A **69**, 063416 (2004).
10. N. Milosevic, P. B. Corkum, and T. Brabec, Phys. Rev. Lett. **92**, 013002 (2004).
11. K. J. Schafer, M. B. Gaarde, A. Heinrich, *et al.*, Phys. Rev. Lett. **92**, 023003 (2004).
12. N. M. Naumova, J. A. Nees, I. V. Sokolov, *et al.*, Phys. Rev. Lett. **92**, 063902 (2004).
13. F. Lindner, G. G. Paulus, H. Walther, *et al.*, Phys. Rev. Lett. **92**, 113001 (2004).
14. A. A. W. Zholents and M. Fawley, Phys. Rev. Lett. **92**, 224801 (2004).
15. P. Abbamonte, K. D. Finkelstein, M. D. Collins, and S. M. Gruner, Phys. Rev. Lett. **92**, 237401 (2004).
16. X. Song, S. Gong, W. Yang, and Z. Xu, Phys. Rev. A **70**, 013817 (2004).

17. J. Mauritsson, P. Johnsson, R. Lopez-Martens, *et al.*, Phys. Rev. A **70**, 021801 (2004).
18. R. A. Bartels, M. M. Murnane, H. C. Kapteyn, *et al.*, Phys. Rev. A **70**, 043404 (2004).
19. Z. Chang, Phys. Rev. A **70**, 043802 (2004).
20. Y. Mairesse, A. De Bohan, L. J. Frasinski, *et al.*, Phys. Rev. Lett. **93**, 163901 (2004).
21. T. Brabec and F. Krausz, Rev. Mod. Phys. **72**, 545 (2000).
22. R. Moshammer, W. Schmitt, J. Ullrich, *et al.*, Phys. Rev. Lett. **79**, 3621 (1997).
23. J. Ullrich, R. Moshammer, H. Kollmus, *et al.*, GSI Scientific Report 2002 (2002), p. 98.
24. A. J. Baltz, Phys. Rev. Lett. **78**, 1231 (1997).
25. A. B. Voitkiv, B. Najjari, and J. Ullrich, J. Phys. B **36**, 2325 (2003).
26. A. J. Baltz, Phys. Rev. A **52**, 4970 (1995).
27. C. Bamber, S. J. Boege, T. Koffas, *et al.*, Phys. Rev. D **60**, 092004 (1999).
28. C. Muller, A. B. Voitkiv, and N. Grun, Phys. Rev. A **67**, 063407 (2003).
29. C. Muller, A. B. Voitkiv, and N. Grun, Phys. Rev. A **70**, 023412 (2004).
30. J. Eichler, Phys. Rev. A **15**, 1856 (1977).
31. A. M. Dykhne and G. L. Yudin, Usp. Fiz. Nauk **125**, 377 (1978) [Sov. Phys. Usp. **21**, 549 (1978)].
32. A. B. Migdal, *Qualitative Methods in Quantum Theory* (Nauka, Moscow, 1975; Benjamin, Reading, Mass., 1977).
33. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory*, 4th ed. (Nauka, Moscow, 1989; Oxford Univ. Press, Oxford, 1980).
34. V. I. Matveev and É. S. Parilis, Usp. Fiz. Nauk **138**, 583 (1982) [Sov. Phys. Usp. **25**, 881 (1982)].
35. V. I. Matveev, Fiz. Élem. Chastits At. Yadra **26**, 780 (1995) [Phys. Part. Nucl. **26**, 329 (1995)].
36. S. E. Harris and A. V. Sokolov, Phys. Rev. Lett. **81**, 2894 (1998).
37. I. P. Christov, M. M. Murnane, and H. C. Kapteyn, Opt. Commun. **148**, 75 (1998).
38. A. V. Sokolov, D. D. Yavuz, and S. E. Harris, Opt. Lett. **24**, 557 (1999).
39. A. E. Kaplan and P. L. Shkolnikov, Phys. Rev. Lett. **88**, 074801 (2002).
40. J. Eichler, Phys. Rep. **193**, 165 (1990).
41. V. I. Matveev, Zh. Éksp. Teor. Fiz. **124**, 1023 (2003) [JETP **97**, 915 (2003)].
42. V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics*, 3rd ed. (Nauka, Moscow, 1989; Pergamon, Oxford, 1982).
43. R. Anholt, Phys. Rev. A **19**, 1004 (1979).
44. C. A. Bertulani and G. Baur, Phys. Rep. **163**, 299 (1988).
45. J. Eichler and W. E. Meyrhoft, *Relativistic Atomic Collisions* (Academic, New York, 1995).
46. D. M. Davidovic, B. L. Moiseivitsch, and P. H. Norrington, J. Phys. B **11**, 847 (1978).
47. D. C. Ionescu and J. Eichler, Phys. Rev. A **48**, 1176 (1993).
48. C. G. Darwin, Proc. R. Soc. London, Ser. A **118**, 654 (1928).
49. E. G. Thrapsaniotis, Phys. Rev. A **70**, 033410 (2004).
50. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 2: *The Classical Theory of Fields*, 7th ed. (Nauka, Moscow, 1988; Pergamon, Oxford, 1975).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 100, No. 6, 2005, pp. 1050–1060.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 6, 2005, pp. 1195–1206.

Electron Acceleration in Quasi-Stationary Electromagnetic Fields during the Self-Channeling of Intense Light Pulses

A. L. Galkin, V. V. Korobkin, M. Yu. Romanovsky, and O. B. Shiryaev

Prokhorov Institute of General Physics, Russian Academy of Sciences, ul. Vavilova 38, Moscow, 119991 Russia

e-mail: galkin@kapella.gpi.ru; obs@kapella.gpi.ru

Received September 14, 2004

Abstract—We theoretically investigate the possibility of electron acceleration during the self-channeled propagation of laser radiation. We consider a new acceleration mechanism associated with the formation of an ion cloud in material (under the ponderomotive force of the laser radiation) that moves together with the laser pulse. We show that the quasi-stationary electric and magnetic fields generated by the moving ion cloud can lead to the acceleration of electrons up to energies of several dozen MeV and to the formation of an electron beam propagating forward coaxially with the laser pulse. The calculated angular distribution of the accelerated electrons is in satisfactory agreement with published experimental results. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Producing directional electron beams with energies of several MeV and durations of several femtoseconds is interesting from a fundamental standpoint and important from the standpoint of possible applications. In recent years, significant advances have been made in this direction using intense short laser pulses. The generation of such beams has been studied theoretically and experimentally in a considerable number of publications (see, e.g., [1–4]).

The generation of electron beams is directly related to the pattern of nonlinear propagation of laser pulses in a material: intense laser radiation ionizes the material and produces quasi-stationary electron-accelerating electromagnetic fields in it. Such beams are generated both when intense radiation propagates in gaseous media and when radiation is focused on the surfaces of solid targets. The beam parameters depend strongly on the target type and basic characteristics of the nonlinear interaction. Some of the questions on the dynamics of intense laser pulses in a material were considered in the monograph [5].

There are several possible electron acceleration mechanisms. Among the most important mechanisms, we note, first, the generation of plasma Langmuir (the so-called wake) waves behind the pulse and the acceleration of electrons in them and, second, the charge displacement by the ponderomotive force of the laser radiation and the acceleration of electrons by the quasi-static electric field that emerges from this displacement. Since these two mechanisms are in a way competing ones, their comparative analysis is needed, which requires studying both mechanisms. The first mechanism was investigated in sufficient detail in earlier publications. The objective of this paper is to study the second mechanism in more detail.

Let us consider the second mechanism, as applied to the regime of self-channeled propagation [6, 7], in which the laser pulse propagates without changing its divergence to a distance much larger than the diffraction length. The charge displacement by the ponderomotive force of the laser radiation is particularly efficient in this regime. A large self-channeling length can be achieved only in gaseous media, and precisely this case is considered here.

There are several physical mechanisms of the change in the refractive index that lead to self-channeling. The most important of them include increase in the electron mass in a relativistic-intensity laser field, the expulsion of electrons by the ponderomotive force from the region of a strong field, change in the refractive index of the plasma in the quasi-stationary magnetic fields produced by the moving uncompensated ion charge and the current of accelerated electrons, possible generation of plasma waves, etc. These mechanisms have been repeatedly discussed in the literature (see, e.g., [5–9]). Here, we do not consider these mechanisms and assume that all of the conditions necessary for self-channeling to take place are satisfied, i.e., the radiation power $P > P_{cr}$. The critical power P_{cr} decreases with increasing electron density. An expression for P_{cr} that is valid in a wide range of electron densities, including the densities near their critical values, is given in [10].

Since the channel diameter is comparable to the laser wavelength in the case of self-channeled propagation, the transverse intensity gradients are large, which, in turn, yields a large ponderomotive force. Electrons are expelled by this force in the radial direction from the region in which intense laser radiation is concentrated, and the expulsion results in the formation of a region of positive uncompensated charge, a kind of a cloud composed of ions (Fig. 1). This cloud (with a charge that is several orders of magnitude larger than the electron charge) moves together with the laser pulse and with its velocity, although the ions themselves remain stationary. At the leading edge of the pulse, ionization takes place and electrons are expelled by the ponderomotive force; after the passage of the trailing edge of the laser pulse, the ions are neutralized by cold electrons from the surrounding plasma. (Under certain conditions, when the pulse duration exceeds the characteristic time determined by the diameter of the channel and the ion density in it, a Coulomb explosion of the channel can take place, but we do not consider this situation here.) A fraction of the electrons expelled from the channel remain near the ion cloud, partially screening it. Therefore, the electrostatic interaction of the ion cloud with an individual electron outside the cloud is determined not by the total charge of the cloud, but by an effective charge whose value generally decreases with increasing distance from the cloud. The electron motion outside the region of action of the ponderomotive force is determined by the quasi-stationary electromagnetic field of the ion cloud and is predominantly longitudinal.

In this paper, we use an approach based on the solution of the Newton equations; these equations are solved for single electrons, and their reverse effect on the motion of the ion cloud is disregarded. The ion cloud is approximated by large particles [11]. Since the ion cloud moves along the axis with a velocity close to the speed of light, our model includes the delay in the charge interaction and is based on the delayed Lienard–Wiechert potentials [12]. Note that the currently used PIC methods do not properly take into account the delay in the interaction of rapidly moving charges.

The model of electron acceleration in the field of the ion cloud developed here allows the formation of an electron beam to be described quantitatively.

2. THE MOTION OF AN ELECTRON IN THE FIELD OF A POINT CHARGE

To ascertain the main features of the electron acceleration by the moving ion cloud, let us first consider the relativistic motion of an electron in the field of a point charge. This problem was considered in the monograph [12], in which the motion of an electron in the field of a stationary point charge was investigated. Generalizing the solutions obtained in [12] to the case of a point charge moving in the laboratory frame of reference with a relativistic velocity is of considerable interest.

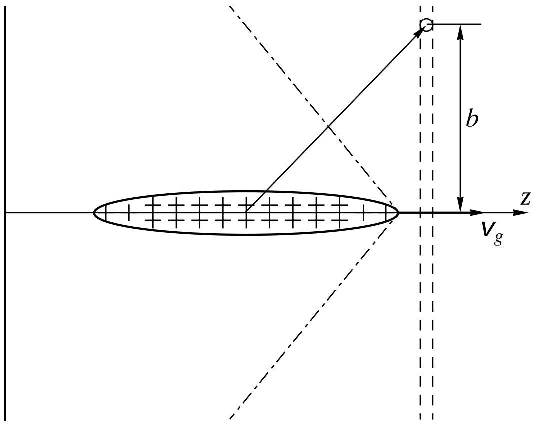


Fig. 1. The formation of a region of positive uncompensated charge, a kind of a cloud composed of ions and moving together with the laser pulse and with its velocity. At the leading edge of the pulse, ionization takes place and electrons are expelled by the ponderomotive force; after the passage of the trailing edge of the laser pulse, the ions are neutralized by cold electrons from the surrounding plasma.

Let us consider the motion of an electron in the field of a positive charge Q. If the mass of the positive charge is much larger than the electron mass, then the change in the energy of the positive charge can be ignored. In this case, the total energy of the electron in the frame of reference in which the positive charge is at rest (the intrinsic frame of reference) is

$$W = c\sqrt{p^2 + m^2c^2} + \alpha/r,$$

where $\alpha = qQ$ is the product of the charges (during the interaction between a positive charge and an electron, $q = -e$ and $\alpha < 0$), $p^2 = M^2/r^2 + p_r^2$, p_r is the radial momentum, and M is a constant angular momentum.

The action as the solution of the Hamilton–Jacobi equation is constructed from general considerations:

$$S = -Wt + M\theta + \int \sqrt{\frac{(W-\alpha/r)^2}{c^2} + \frac{M^2}{r^2} - m^2c^2}\, dr.$$

The trajectory can be obtained from the equation $\partial S/\partial M = 0$. It can be written in analytical form: $\theta = \theta(r)$. The inverse function $r = r(\theta)$ also has an analytical expression [12] that contains a dimensional constant,

$$r_0 = \frac{\alpha^2 - c^2M^2}{c\sqrt{M^2W^2 + m^2c^2(\alpha^2 - c^2M^2)}}.$$

The condition for the capture of an electron by a positive charge is $cM < |\alpha|$. In this case, the trajectory in polar coordinates is a convergent spiral whose θ dependence is given by the hyperbolic cosine

$$\cosh(\theta\sqrt{(eQ/cM)^2 - 1}),$$

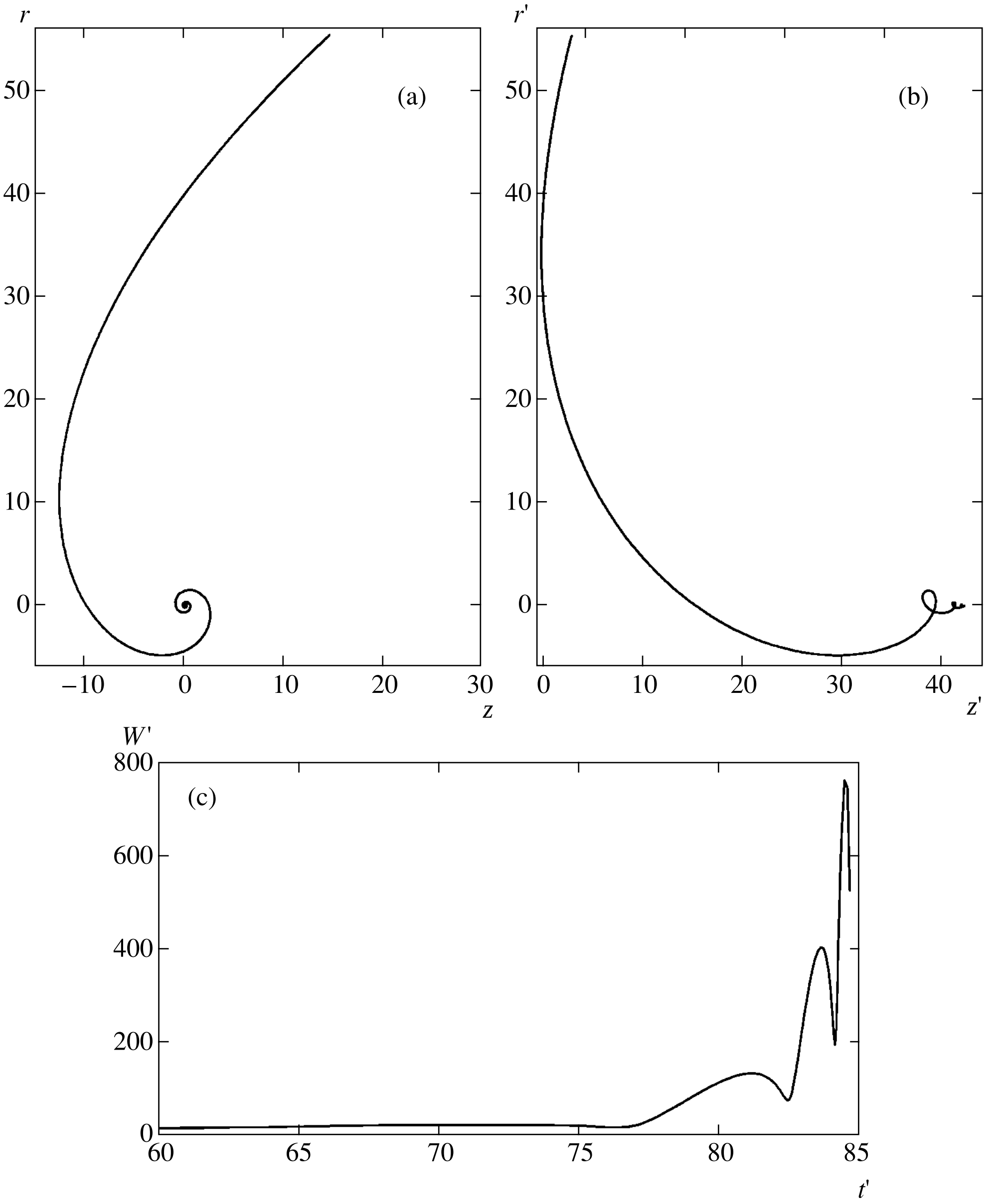


Fig. 2. Electron trajectories in the (a) intrinsic and (b) laboratory frames of reference for the same parameters and $V = c/2$; (c) kinetic energy versus time in the laboratory frame. All parameters are given in dimensionless form: the space coordinates are in units of r_0, the time is in units of r_0/c, and the energy is in units of mc^2.

so the number of complete spiral turns from $\theta = 0$ to the fall to the center can be estimated as

$$\frac{1}{2\pi\sqrt{(eQ/cM)^2 - 1}}.$$

At the parameters typical of the problem under consideration, the electron can fall rapidly to the positive charge; in this case, the number of complete turns is much less than 1.

The time dependence $t = t(r)$ can be derived from the condition $\partial S/\partial W = 0$. This function has an analytical expression, while the inverse function $r = r(t)$ has no analytical expression; therefore, our subsequent analysis is based on numerical calculations. The time dependences of the Cartesian coordinates and the velocity components can be numerically derived from the functions $\theta(r)$ and $t(r)$. The values of t, θ, x, y, $\partial r/\partial t$, $\partial\theta/\partial t$, v_x, v_y, and the electron kinetic energy are determined sequentially on a uniform r coordinate grid.

The Lorentz transformation allows us to numerically derive similar dependences, including the electron kinetic energy, in the laboratory frame of reference, in which the positive charge moves along the z axis with a constant velocity V.

Figure 2a shows an example of the trajectory along which the electron falls to the center in the intrinsic frame of reference. Going to the laboratory frame of reference leads to a significant change of the trajectory in the shape of an open spiral. As an example, Fig. 2b shows the shape of the trajectory in the laboratory frame of reference for the same parameters and $V = c/2$ as those in Fig. 2a.

An analysis of the electron motion leads us to two conclusions. First, during its motion, the electron can overtake the positive charge moving with a constant velocity. Second, the electron kinetic energy oscillates: it has maxima and minima attributable to the overtaking of the positive charge during the motion in one direction and the lagging behind it during the motion in the opposite direction (in contrast to the motion in the frame of reference in which the positive charge is at rest, where the electron kinetic energy monotonically increases). In this case, the electron kinetic energy can significantly exceed $W^* = mc^2/\sqrt{1 - V^2/c^2}$. The time dependence of the kinetic energy is shown in Fig. 2c. In our calculations, we used $Q/e \approx 2.5 \times 10^{10}$ and $M = 0.99Qe/c$.

In the case under consideration, the velocity of the positive charge is determined by the velocity of the ionization wave, which is equal to the group velocity of the laser pulse in the medium. The Coulomb interaction of the electrons with the positive charge allows them to reach velocities much higher than the velocity of the ionization wave.

3. BASIC EQUATIONS DESCRIBING THE PROPAGATION OF A LASER PULSE

3.1. The Physical Picture of the Nonlinear Propagation of a Laser Pulse

An intense laser pulse propagating through a medium interacts with the medium and greatly changes its properties. This interaction manifests itself in several aspects. The pulse ionizes the medium even at its leading edge, and the bulk of the pulse propagates in the plasma with the group velocity $V_g < c$ determined by the plasma parameters. We assume that the radiation propagates in hydrogen, the produced plasma is hydrogenic, and the initial electron, n_e, and ion, n_i, densities are equal, $n_e = n_i = n$. When the radiation propagates in other gases, multiple ionization takes place, $n_e = Zn_i$, and the total charge of the ion cloud is a factor of Z larger. Note that the medium is ionized and the plasma is produced mainly inside the laser pulse. However, due to the plasma radiation in the short-wavelength spectral range, a plasma halo in which the electron density decreases rapidly with increasing distance from the axis is formed around the region occupied by the laser pulse. Inside the beam, the plasma is in the field of intense laser radiation, and the plasma parameters, including its refractive index, are determined by this radiation. In turn, a change in the refractive index modifies the pattern of propagation of the radiation itself. It is well known [7] that the pulse propagates in the regime of self-channeling if its total power exceeds its critical value. In this regime, two mechanisms of the change in the plasma refractive index play a major role: the relativistic increase in the electron mass and the expulsion of electrons by the ponderomotive force from the region of a strong field. Both mechanisms cause the plasma refractive index in the region of a strong field to increase, thereby ensuring the necessary conditions for self-channeling. As a result, only ions are left inside the channel, and an ion cloud (coinciding with the region of action of the ponderomotive force) is formed; this cloud effectively moves with the group velocity of the laser pulse. The ions remain stationary, while the positive charge at the trailing edge of the laser pulse is neutralized by cold electrons from the cloud-surrounding plasma. The source of this plasma, which emerges only in the immediate vicinity of the ion cloud, is the radiation from the ion cloud, including its X-ray radiation. Note that the electron density in the produced plasma decreases rapidly with increasing distance from the beam axis.

The sizes of the ion cloud are determined by the charge displacement under the ponderomotive force and by the self-consistent collective motion of electrons (the mobility of ions is low) in the emerging quasi-static electric field. In general, the boundaries of the ion cloud are blurred. In this paper, the cloud diameter a is assumed to be equal to the diameter of the region of a strong field at the center of the laser pulse. The cloud length L was chosen from the condition that the intensity on the axis inside the cloud was higher than $I_{\min} = 3 \times 10^{17}$ W cm^{-2}. This choice is dictated by the fact that at such intensities, the material is completely ionized and the electrons are expelled from the channel by the ponderomotive force. The shape of the ion cloud used in the model is a cylinder with a total volume of $\pi a^2 L$. When considering the Coulomb interaction of the ion cloud with an electron outside it, we assumed that the cloud had an effective charge Q; the latter was obtained by multiplying the total charge of the ions by a corrective coefficient that included many factors. First, the diameter of the actual cloud slightly decreases with increasing distance from the center of the laser pulse; second, as was mentioned above, the positive charge of the cloud is partially screened by the electrons captured by the cloud. In this paper, this corrective coefficient is close to 1/2. The latter value was obtained by analyzing the trajectories of the electrons moving in the field of the ion cloud.

The moving ion cloud produces quasi-stationary electric and magnetic fields in its vicinity. Since the cloud velocity is close to the speed of light, the delayed Lienard–Wiechert potentials must be used to describe these fields. Note that the skin depth for these fields is much larger than the size of the plasma region under consideration; therefore, the plasma is transparent for them. Since $\varepsilon = \mu = 1$ in the original gas, we conclude that the velocity of propagation of these fields is equal to *c*.

The electrons surrounding the ion cloud can be divided into two groups. The electrons that were initially formed inside the ion cloud constitute the first group. The electrons of this group are initially accelerated inside the cloud by the ponderomotive force to high energies and subsequently interact with the positive charge of the ion cloud. The initial acceleration significantly affects their trajectories. The electrons formed near the cloud, but outside the region of action of the ponderomotive force constitute the second group. Immediately after their appearance, the electrons of the second group begin to interact with the fields produced by the ion cloud.

Let us consider the passage of the ion cloud through a thin layer of material located perpendicularly to the direction of its motion. Initially, the material near the axis is ionized by the laser pulse; subsequently, the X-ray radiation propagating from the cloud also ionizes the material at a certain distance from the axis. A peculiar ionization wave propagates through this thin layer; as a result, electrons appear at some impact parameter *b* and begin their motion attributable to the interaction with the fields of the ion cloud with a delay relative to the passage of the leading edge of the pulse through the chosen layer. It can be assumed that a peculiar “light cone” with a vertex half-angle $\varphi = \arctan(V_g/c)$ propagates together with the leading edge of the pulse, and that this cone gives rise to an electron and turns on the electromagnetic action on it from the ion cloud. The delay is approximately equal to b/c.

We assume that at time $t = 0$, the thin layer under consideration is far from the peak of the laser pulse, and the distance to it is $100a$ in the longitudinal direction. Such a large distance allows the initial formation of the ion cloud and the appearance of the fields as the pulse enters into the gas to be not considered.

3.2. The Equation of Motion for Electrons

Three forces act on an electron: the ponderomotive force produced by laser intensity gradients and the two forces governed by the electric and magnetic fields of the ion cloud. Suppose that the laser radiation is a short pulse of duration τ and radius a that propagates along the z axis and has an intensity distribution $I(r, z, t)$.

The ponderomotive force acting on an electron in a relativistic-intensity field is

$$\mathbf{F}_p = -\frac{I_r}{cn_{e,\,\mathrm{cr}}}\nabla\left(1+\frac{I}{I_r}\right)^{1/2}, \tag{1}$$

where $I_r = 2.75 \times 10^{18}(1/\lambda[\mu\mathrm{m}])^2$ W cm^{-2} is the relativistic intensity, and $n_{e,\,\mathrm{cr}}$ is the critical plasma electron density. Expression (1) is valid for radiation with circular polarization, but can also be approximately used for linear polarization. In the nonrelativistic case, expression (1) is

$$\mathbf{F}_p = -\frac{1}{2cn_{e,\,\mathrm{cr}}}\nabla I.$$

In general, ∇I has both radial and longitudinal components; as a result, the ponderomotive force also has the same components, F_{pr} and F_{pz}.

Let us assume that the intensity distribution in the beam is

$$I(r, z, t) = I_0 f_1(r) f_2(z, t), \tag{2}$$

where the radial distribution is given by

$$f_1(r) = \exp\left[-\ln 2\left(\frac{r}{a}\right)^{\gamma}\right], \tag{3}$$

and the time dependence is

$$f_2(z, t) = \exp\left[-4\ln 2\left(\frac{z - z_0 - V_g t}{\tau V_g}\right)^2\right]. \tag{4}$$

The parameter γ in expression (3) determines the shape of the transverse intensity distribution of the laser pulse. For self-channeled propagation, this parameter is larger than 2. In expression (4), z_0 denotes the initial coordinate of the laser pulse intensity peak and the center of the ion cloud. For plasma, the pulse group velocity is $V_g = (d\omega/dk)_{\omega_0} = cN$, where N is the refractive index of the plasma.

Note that dependence (4) on the longitudinal coordinate implies the passage to an axially symmetric model, in contrast to the cylindrically symmetric model used previously [13]. In specific calculations, the intensity $I(r, z, t)$ was taken to be zero if its value defined by formula (2) was lower than a certain minimum value, $I \le I_{\min}$.

The distance between the leading and trailing edges of the pulse can be determined from (4):

$$L = 2\left[-\frac{\ln(I_{\min}/I_0)}{4\ln 2}\right]^{1/2}\tau V_g. \tag{5}$$

The charge of the ion cloud composed of hydrogen ions with density n_i is

$$Q = \frac{1}{2}\pi a^2 L n_i e, \tag{6}$$

where e is the elementary charge.

Denote the electric and magnetic fields produced by the moving ion cloud in the surrounding space by **E** and **H**, respectively. In this case, the equation of motion for an electron is

$$\frac{d}{dt}\frac{m_e\mathbf{V}}{\sqrt{1-V^2/c^2}} = -e\mathbf{E} - \frac{e}{c}\mathbf{V}\times\mathbf{H} + \mathbf{F}_p, \qquad (7)$$

where **V** is the electron velocity, $V = |\mathbf{V}|$.

The fields produced by a point charge Q moving with a constant velocity V_g is defined by the expressions (the delayed Lienard–Wiechert potentials)

$$\begin{aligned} \mathbf{E} &= \frac{Q\mathbf{R}}{R^3}\frac{1-V_g^2/c^2}{(1-V_g^2\sin^2\theta/c^2)^{3/2}}, \\ \mathbf{H} &= \frac{QV_g[\mathbf{n}\times\mathbf{R}]}{R^3c}\frac{1-V_g^2/c^2}{(1-V_g^2\sin^2\theta/c^2)^{3/2}}, \end{aligned} \qquad (8)$$

where **n** is a unit vector along the z axis, θ is the angle between the z axis and the radius vector **R** of the electron, $R = |\mathbf{R}|$

At time $t = 0$, the laser pulse peak has the coordinate z_0 and moves along the z axis with the velocity V_g (the center of the ion cloud has the same coordinate and velocity), the electron coordinates are $r(0) = b$ and $z(0) = 0$, and the electrons are initially at rest.

The componentwise representation of (7) in the dimensionless variables

$$t_1 = \frac{ct}{a}, \quad v_r = \frac{V_r}{c}, \quad v_z = \frac{V_z}{c}, \quad v_g = \frac{V_g}{c}$$

is

$$\begin{aligned} \frac{d}{dt_1}\frac{v_z}{\sqrt{1-v^2}} &= -A\frac{1-v_g^2}{(1-v_g^2\sin^2\theta)^{3/2}} \\ &\quad\times\frac{z_1+v_gv_rr_1}{R_1^3} + F_{pz}\frac{a}{m_ec^2}, \end{aligned} \qquad (9)$$

$$\begin{aligned} \frac{d}{dt_1}\frac{v_r}{\sqrt{1-v^2}} &= -A\frac{1-v_g^2}{(1-v_g^2\sin^2\theta)^{3/2}} \\ &\quad\times\frac{r_1(1-v_gv_z)}{R_1^3} + F_{pr}\frac{a}{m_ec^2}, \end{aligned} \qquad (10)$$

where

$$z_1 = \frac{z(t)-z_0-V_gt}{a}, \quad r_1 = \frac{r(t)}{a}, \quad R_1^2 = z_1^2 + r_1^2,$$

$$A = \frac{Qe}{am_ec^2}, \quad \sin\theta = \frac{r_1}{R_1}.$$

The terms with the factor A on the right-hand sides of Eqs. (9) and (10) describe the Coulomb and Lorentz forces (the latter contain the components v_gv_r and v_gv_z); the quantities r_1, z_1, and R_1 are the components and magnitude of the radius vector of the electron relative to the point ion in units of a.

In the case under consideration, the length of the ion cloud is much larger than its diameter. Therefore, we must either generalize Eqs. (9) and (10) to the case of a distributed charge or approximate the ion cloud by a system of point charges with an adjustment of the field of these charges. We used the second alternative, in which the cloud was modeled by a system of $K = 50$ identical point charges uniformly arranged on the axis. Three different cases can be distinguished, depending on the distance of the accelerated electron from the axis. At a large distance ($R_1 > 1$), the value of each point charge was taken to be $q = Q/K$, where K is the number of point charges. If, however, the electron penetrates into the ion cloud to a distance $1 \ge R_1 \ge 0.1$, then its motion in this region may be considered as the motion in the field of a cylindrical charge. The influence of the ions located in the outer (with respect to the charge) regions of the cylinder, i.e., at distances larger than R_1, can be disregarded. This approximation is justified everywhere, except the small regions at a distance of the order of the radius from the cylinder ends. In this case, the cylindrical charge can be approximated by a system of point charges with the value of each point charge $q = R_1^2 Q/K$. When the electron is located at distances $R_1 \le 0.1$, the approximation of point charges arranged on the axis is no longer valid, since the fields on the cloud axis have a singularity in this approximation. To make the passage of electrons through the region near the axis possible (which may occasionally be required), this singularity must be removed. To this end, we assumed in the region $R_1 \le 0.1$ that the electron interacts with one point charge with $q = R_1^3 Q/K$.

Thus, the charge is $q = \chi Q/K$, where

$$\chi = \begin{cases} 1, & R_1 \ge 1, \\ R_1^2, & 0.1 \le R_1 \le 1, \\ R_1^3, & R_1 \le 0.1. \end{cases}$$

Formally, allowance for the finite cloud sizes in (9) and (10) consists in going to the summation over all charges q_j on the right hand sides of these equations ($j = 1, \ldots, K$). In this case, the quantities r_1, z_1, R_1, and θ in these equations are replaced with r_{1j}, z_{1j}, R_{1j}, and θ_j. The value of $j^* = (K+1)/2$ corresponds to the center of the ion cloud. Below, we use $z_1^* = z_{1j^*}$ to denote the electron coordinate relative to the center of the ion cloud.

The presence of $\sin\theta$ in Eqs. (9) and (10) reflects the inclusion of the delay in the electromagnetic action. In addition to this delay, there is also a time factor. It stems from the fact that the ion cloud emerges and forms at a finite distance from the electron under consideration; therefore, a regime described by the equations for the delayed potentials is reached. However, analysis shows that this effect is insignificant in the conditions under consideration.

3.3. The Angular Distribution of Accelerated Electrons

As follows from Eqs. (7) and (8), the electron moves in the plane passing through the z axis and the point of its initial location. Let us consider its motion in a two-dimensional Cartesian coordinate system located on this plane.

The solution of the initial system defines the velocity components, coordinates, and kinetic energy as functions of the impact parameter b and time t. The angle between the electron velocity direction and the axis after its interaction with the ion cloud was taken as the exit angle. This angle is a single-valued function of the impact parameter b, $\theta_e = \theta_e(b)$. Since the electron can cross the z axis as it moves, the exit angle of the accelerated electron, θ_e, can be both positive and negative, depending on the specific impact parameter b. Note also that during its motion, the electron can be multiply reflected from the region with a large ponderomotive force; its trajectory can be complex. Therefore, the exit angle θ_e can have equal values for different b; i.e., the inverse function $b(\theta_e)$ is generally not single-valued.

In the problem under consideration, the electrons are located in both the upper and lower half-planes. Formally, as a result of this, both positive and negative values of b should be taken into account when calculating the angular distribution. Due to the symmetry properties of the electric and magnetic field distributions, the electron trajectories at the initial values of b and $-b$ are mirror-symmetric relative to the z axis, which ultimately yields $\theta_e(-b) = -\theta_e(b)$.

To calculate the angular distribution of the accelerated electrons, we must go from the two-dimensional Cartesian coordinate system in which the electron trajectories were analyzed to a cylindrical coordinate system. The sought-for angular distribution must be axially symmetric relative to the z axis.

Given the axial symmetry, the total number of electrons in a cylindrical layer of unit length with width db at distance b is

$$P(b)db = 2\pi n_e(b)bdb. \tag{11}$$

On the single-valued segments of the function $b(\theta_e)$, it transforms into the angular electron distribution

$$P(\theta_e)d\theta_e = 2\pi n_e(b)b\left|\frac{db}{d\theta_e}\right|d\theta_e. \tag{12}$$

Since $b(\theta_e)$ is not a single-valued function, expression (12) must be summed over its single-valued branches in a given b range.

The calculation of (12) can be simplified appreciably if we first average it over the positive and negative values of b.

Since $\theta_e(-b) = -\theta_e(b)$, the averaging of the distribution is achieved by going to the quantity $[P(\theta_e) + P(-\theta_e)]/2$, which should be considered only at positive θ_e. After this procedure, the angle θ_e acquires the meaning of an azimuthal angle in the angular distribution of the accelerated electrons.

The following should be taken into account when calculating the angular distribution. In expressions (11) and (12), the electron density distribution $n_e(b)$ depends on the distance b to the axis, since the ionization of the material is determined by the intensity of the short-wavelength radiation from the ion cloud. This dependence is different in three different regions. In the first region that extends from the channel axis to a distance of $(1–2)a$ from the surface of the ion cloud, i.e., for $0 \le b \le 3$, the intensity of this radiation is approximately constant. Therefore, the electron density is also constant. The second region extends from the boundary of the first region to the distance L from the axis (L is the cloud length). In this region, the ion cloud can be treated as an extended source of X-ray radiation. Therefore, it should be assumed that its intensity (and, accordingly, the electron density) in the second region varies as $1/b$. In the third region beginning from $b \approx L$, the ion cloud must be treated as a point source of X-ray radiation. Therefore, it should be assumed that its intensity (and, accordingly, the electron density) in the third region varies as $1/b^2$ or even faster. The latter stems from the fact that during ionization, the X-ray radiation is absorbed and its intensity decreases.

It follows from the foregoing that the contribution of various impact parameters to the formation of an electron beam can be determined by taking into account the corresponding dependences $n_e(b)$.

4. DISCUSSION OF THE RESULTS OF NUMERICAL CALCULATIONS

4.1. The Capture of an Electron by a Point Ion

To test the code, we computed the capture of an electron by a positive point charge Q.

For $n/n_{e,\,\mathrm{cr}} = 0.1$ and $V_g/c = 0.95$, the electron has the angular momentum

$$M = \frac{0.95cm_eb}{\sqrt{1-0.95^2}},$$

and the condition $M < M_{\mathrm{cr}} = |\alpha|/c$ transforms to the condition $b < 13.9a$.

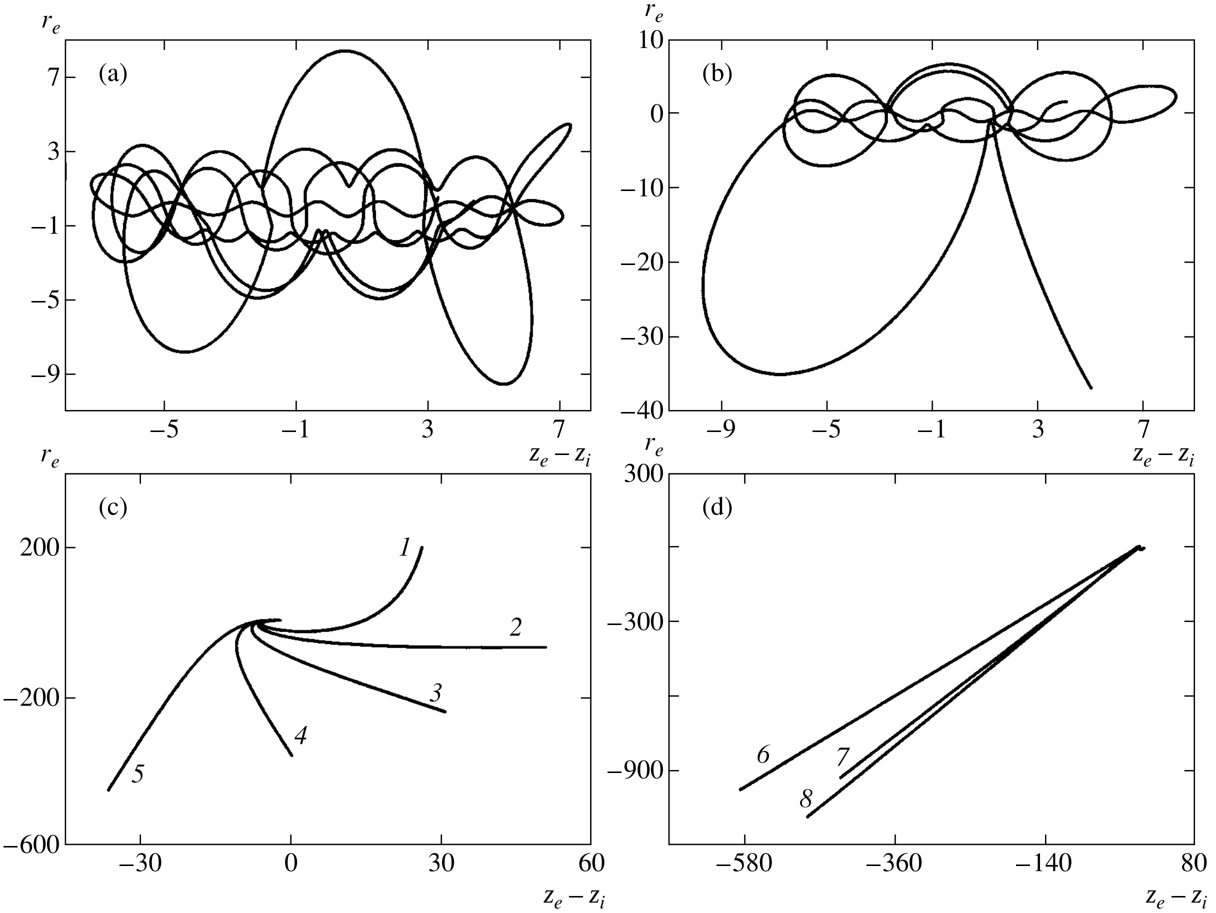


Fig. 3. Electron trajectories for $I_0 = 10^{19}$ W cm^{-2} and $n/n_{e,\,\mathrm{cr}} = 0.075$ (a–c) and 0.04 (d) at $b = 1.925$ (a), 1.95 (b), 8.5 (*1*), 8.8 (*2*), 9.1 (*3*), 9.4 (*4*), 9.7 (*5*), 1.95 (*6*), 2.375 (*7*), and 8.5 (*8*).

The results of our numerical calculations are in close agreement with the analytical results presented in Section 2.

4.2. Specifying the Parameters of the Problem

In our calculations, we assumed that the laser radiation had a wavelength of 1.06 μm at the pulse duration $\tau = 50$ fs ($n_{e,\,\mathrm{cr}} = 9.93 \times 10^{20}$ cm^{-3}, $I_r = 2.45 \times 10^{18}$ W cm^{-2}) with a hyper-Gaussian (3) radial intensity distribution with $\gamma = 6$ and a Gaussian (4) longitudinal and time dependence $f_2(z, t)$. The calculations were performed for $a = 3$ μm, initial particle densities $n/n_{e,\,\mathrm{cr}}$ in the range 0.03–0.5, and peak intensities in the range 4×10^{18}–4×10^{20} W cm^{-2}. The value of z_0 was taken to be $-100a$. We assumed that the laser pulse entered the medium at time $t = 0$ and analyzed the electron motion over a period of $t = 2000a/c$.

4.3. Results of Numerical Calculations

By integrating the system of equations (9) and (10), we determined the time profiles of both (longitudinal and radial) coordinates, both velocities, and the kinetic energy of the electron with various initial impact parameters *b*.

The configurations of the electric and magnetic fields in the laboratory frame of reference produced by the moving cloud differ markedly from those for a point charge. (Note that the magnetic field strength in the frame of reference associated with the moving ion cloud is $H \equiv 0$.) They are more elongated along the *z* axis. The characteristic electric field strength is $E_0 \approx 5 \times 10^{13}$ V m^{-1}, while the characteristic magnetic field strength is $H_0 \approx 2 \times 10^{10}$ A m^{-1}.

Figure 3 shows the electron trajectories for $I_0 = 10^{19}$ W cm^{-2} and two values of $n/n_{e,\,\mathrm{cr}}$. For all of the trajectories shown in Fig. 3, $z_e - z_i$ associated with the motion of the ion cloud and r_e are used as the longitudinal and transverse variables, respectively.

Analysis of the results reveals two types of trajectories that correspond to two ranges of the parameter *b* separated by a critical value b_{cr}. Just as in the case of a point ion, the electron passes by the cloud as it moves at $b > b_{\mathrm{cr}}$ and is captured by the cloud at $b < b_{\mathrm{cr}}$. In the

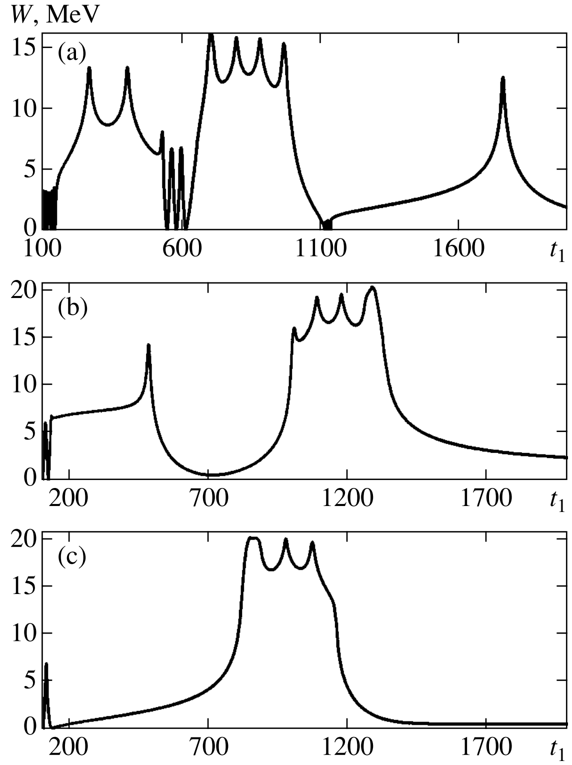


Fig. 4. Electron kinetic energy versus time for I_0 = 10^{19} W cm^{-2} and $n_0/n_{e,\,\mathrm{cr}}$ = 0.075 at b = 1.95 (a), 5.525 (b), and 6.825 (c).

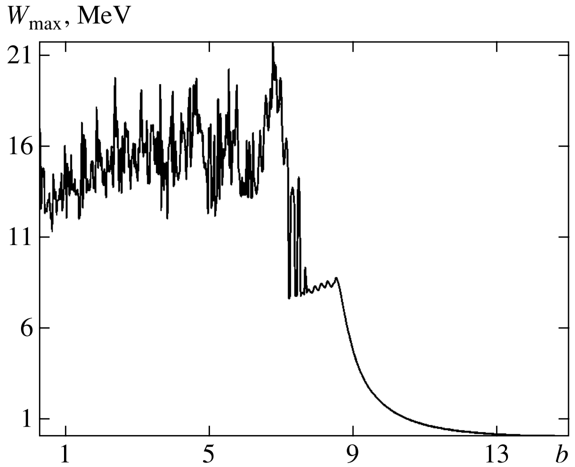


Fig. 5. Maximum electron energy $W_{\max}$ [MeV] versus impact parameter b in the time interval t_1 = 0–2000.

trajectories of the first type shown in Figs. 3a and 3b, the electron is captured by the ion cloud. In the capture regime, the electron is in a relatively small vicinity of the cloud. At high intensities, it can be multiply reflected from the region of action of the ponderomotive force without going into it, while at low intensities, it can go into the region of action of the ponderomotive force while crossing the z axis. The angular momentum **M** of an electron in a centrally symmetric field is known to be conserved, and the capture condition is $|\mathbf{M}| \le M_{\mathrm{cr}}$. During the interaction with a distributed charge (cloud), the angular momentum is not conserved. Moreover, during the reflection from the region of action of the ponderomotive force, the angular momentum of the electron changes sharply; therefore, the capture is not continuous, and the electron escapes from the capture regime and recedes from the cloud in a certain time (Fig. 3b).

Figure 4 shows the time dependences of the electron kinetic energy for $I_0 = 10^{19}$ W cm^{-2} and $n/n_{e,\,\mathrm{cr}} = 0.075$ for three impact parameters: b = 1.95 (Fig. 4a), 5.525 (Fig. 4b), and 6.825 (Fig. 4c). In Fig. 4a, the electron is in the capture regime; in Figs. 4b and 4c, the electron escapes from the capture regime at a different time. Analysis of these curves leads us to the following conclusions. In the capture regime, the energy oscillates with time, reaching its maximum when the electron moves around the ion cloud is at the minimum distance from the cloud and only begins to be decelerated by the ponderomotive force. The electron has a minimum energy of ~3–4 MeV at the maximum distance from the ion cloud. The electron also has approximately the same energy after its escape from the capture regime far from the cloud.

As $n/n_{e,\,\mathrm{cr}}$ decreases, the total charge of the ion cloud decreases, causing the Coulomb force to decrease. In this case, the regime of electron motion around the ion cloud changes: the capture regime disappears, and the electron moves around the ion cloud in a smooth trajectory. The change of regimes occurs at $n/n_{e,\,\mathrm{cr}} = 0.05$.

The maximum energy achievable in the capture regime also depends on the density. It is 20 MeV at $n/n_{e,\,\mathrm{cr}} = 0.075$ and reaches 140 MeV when the density increases to $n/n_{e,\,\mathrm{cr}} = 0.5$.

In Fig. 5, the maximum electron energy $W_{\max}$ in the capture regime is plotted against the impact parameter b. It is interesting to compare this quantity with the energy of the electron oscillations in the laser radiation field, which is 0.6 MeV for the intensity $I_0 = 10^{19}$ W cm^{-2} under consideration. (This is the energy of the electrons with $b \le 0.2a$ after their escape from the region of action of the ponderomotive force.) It thus follows that the electron acquires an energy higher than 0.6 MeV during the Coulomb interaction with the cloud.

4.4. The Angular Distribution of Accelerated Electrons

The accelerated electrons escaped from the capture regime fly at an angle $\theta_e(b)$ to the axis; this angle has a random value in some interval (Fig. 6a). Electrons with different impact parameters can fly at the same angles to the axis. Figure 6b shows the angular electron distribution calculated using the technique described above.

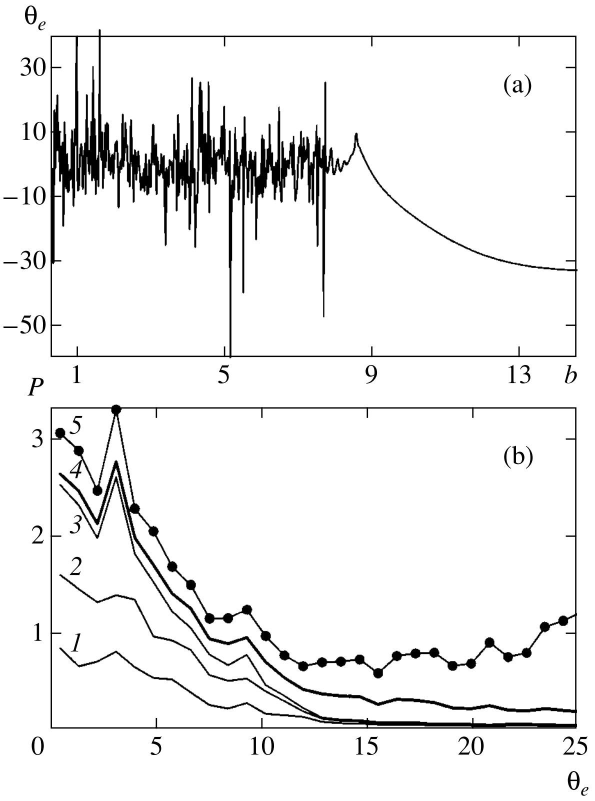


Fig. 6. (a) Electron exit angle θ_e versus impact parameter b; (b) the angular electron distribution. Curve *4* represents the resulting angular distribution for the interval $0.1a \le b \le 12a$, curve *5* represents the resulting angular distribution under the assumption of $n_e(b)$ = const for the interval $0.1a \le b \le 12a$. Curves *1*, *2*, and *3* were obtained for the intervals $0.1a \le b \le 5a$, $0.1a \le b \le 7a$, and $0.1a \le b \le 8.8a$, respectively.

Curve *4* represents the resulting angular distribution. Curve *5* is shown for comparison. Curves *1*, *2*, and *3* illustrate the contributions of various impact parameters to the ultimate angular distribution.

A comparison of curve *4* with experimental results is of considerable interest. In this case, the experimental results obtained under the self-channeling conditions must be used. Curve *4* is in satisfactory agreement with the experimental curve obtained in [2] during the propagation of laser radiation in gases. Note that it is improper to compare the above computational data with the results of the experiments in which laser radiation was focused on the surface of a solid target.

The time t_c during which the electron is in the capture regime depends on the impact parameter b. This dependence is random, and t_c can change greatly as b changes only slightly. Nevertheless, our calculations show that for the thin layer under consideration, we can introduce a mean value of t_c^* that corresponds to the condition under which the fraction of the electrons left in the capture regime is $1/e$ of their total number in this layer. The quantity $v_g t_c^*$ gives the mean free path of an electron in the captured state. It can be assumed that all electrons escape from the capture regime after the time t_c^* and subsequently propagate in their own trajectories without interacting with the field of the ion cloud. Next, let us consider two cases. In the first case, the total length of the nonlinear medium is $L_{nl} \gg v_g t_c^*$, all thin layers are equivalent, and the angular electron distribution at the exit from the nonlinear medium is similar to that after one thin layer (without the additional scattering of electrons as they propagate in the medium). In the second case, the length of the nonlinear medium is $L_{nl} \le v_g t_c^*$, the thin layers are not equivalent, and the integration must be performed over the entire length of the nonlinear medium when calculating the beam parameters at the exit.

Both the maximum energy during the oscillations in the capture regime and the energy after the recession from the cloud are of considerable interest from the standpoint of possible applications. The former case is important, for example, for the possible use of oscillating electrons to collisionally excite the heavy particles near the ion cloud and to produce inversion in the X-ray range. The latter case deals with the maximum energy of the electrons in the electron beam produced through the passage of an ultrashort pulse through material.

5. CONCLUSIONS

We investigated the possibility of electron acceleration during the self-channeled propagation of laser radiation.

In this regime, an ion cloud moving together with the laser pulse is formed in the material. Our analysis is based on studying the trajectories of single electrons under the ponderomotive force of a laser pulse and the quasi-stationary electric and magnetic fields generated by the moving ion cloud.

Our calculations lead us to the following conclusions.

(1) The maximum energy of the electron, $W_{\max}$, that it reaches during its motion along the trajectory increases with density and depends weakly on intensity.

(2) The maximum output energy of the electron, W_{out} (the electron energy after the propagation of a laser pulse in a nonlinear medium of finite length), increases with density and depends weakly on intensity.

(3) In all regimes, $W_{\max} \ge W_{\text{out}}$.

(4) Two electron acceleration regimes are observed at fixed intensity. At low density, no electrons are captured by the ion cloud and their energies $W_{\max}$ and W_{out} lie in the range of several MeV. At high density, electrons are captured by the cloud, and the energies $W_{\max}$ and W_{out} can reach 100 MeV or more.

(5) The angular distribution of the beam of accelerated electrons is in satisfactory agreement with published experimental results.

Note that there is no plasma, but only a charged ion cloud inside the laser pulse. Thus, no plasma waves can be generated in it. After the passage of the laser pulse through a layer of material, the tail of the ion cloud is rapidly neutralized by the surrounding cold electrons. Under such conditions, the possibility of plasma wave generation in the wake of the laser pulse is not obvious. The electron acceleration mechanism suggested in this paper, in fact, does not require invoking the plasma wave mechanism. The electromagnetic field of the ion cloud traveling together with the laser pulse acts as the wakefield considered previously [9, 14, 15].

ACKNOWLEDGMENTS

This work was supported in part by the Russian Foundation for Basic Research (project nos. 02-01-00744 and 04-02-17259).

REFERENCES

1. T. Tajima and J. Dawson, Phys. Rev. Lett. **43**, 267 (1979).
2. R. Wagner, S.-Y. Chen, A. Maksimchuk, and D. Umstadter, Phys. Rev. Lett. **78**, 3125 (1997).
3. F. Amiranoff, S. Baton, D. Bernard, *et al.*, Phys. Rev. Lett. **81**, 995 (1998).
4. Y. Li, J. Zhang, Z. M. Sheng, *et al.*, Phys. Rev. E **69**, 036405 (2004).
5. A. F. Aleksandrov and A. A. Rukhadze, *Lectures on Electrodynamics of Plasma-like Media. Nonequilibrium Media* (Fiz. Fakul. Mosk. Gos. Univ., Moscow, 2002) [in Russian].
6. A. B. Borisov, A. V. Borovskiy, O. B. Shiryaev, *et al.*, Phys. Rev. A **45**, 5830 (1992).
7. A. B. Borisov, A. V. Borovskiy, V. V. Korobkin, *et al.*, Phys. Rev. Lett. **68**, 2309 (1992).
8. D. Umstadter, J. Phys. D: Appl. Phys. **36**, R151 (2003).
9. A. Pukhov, Rep. Prog. Phys. **66**, 47 (2003).
10. B. Shen and M. Y. Yu, Phys. Rev. E **68**, 026501 (2003).
11. V. G. D'yachenko and V. S. Imshennik, Zh. Éksp. Teor. Fiz. **56**, 1766 (1969) [Sov. Phys. JETP **29**, 947 (1969)].
12. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 7th ed. (Nauka, Moscow, 1978; Pergamon, Oxford, 1975).
13. A. L. Galkin, V. V. Korobkin, M. Yu. Romanovskiĭ, and O. B. Shiryaev, Kratk. Soobshch. Fiz., No. 1, 39 (2002).
14. L. M. Gorbunov and V. I. Kirsanov, Zh. Éksp. Teor. Fiz. **93**, 509 (1987) [Sov. Phys. JETP **66**, 290 (1987)].
15. S. V. Bulanov, T. Zh. Esirkepov, M. Lontane, *et al.*, Phys. Rev. Lett. **76**, 3562 (1996).

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 100, No. 6, 2005, pp. 1061–1068.
Translated from Zhurnal Éksperimental’noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 6, 2005, pp. 1207–1215.
Original Russian Text Copyright © 2005 by Shutyi, Sementsov.

ATOMS, MOLECULES, OPTICS

Quasi-Solitons of Magnetostatic Waves in a Two-Layer Structure under Single-Mode Excitation

A. M. Shutyi and D. I. Sementsov
Ulyanovsk State University, Ulyanovsk, 432970 Russia
e-mail: shuty@mail.ru; sementsovdi@ulsu.ru

Received September 22, 2004

Abstract—We investigate the quasi-solitons of the envelope of forward body magnetostatic waves in a structure that consists of two magnetically coupled films separated by a nonmagnetic interlayer under the initial excitation of a pulse in one of them. We have found that as the pulse propagates, its energy is transferred from one film to the other through intermode coupling; as a result, “pulsating” quasi-solitons of the coupled modes of the two films emerge. We show that the amplitude of each of the mode pulses at the exit from the waveguide can be regulated over a wide range by varying the magnetizing field. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

In recent years, the soliton regimes of propagation of waves of various natures have attracted the rapt attention of researchers. Apart from optical solitons [1], solitons in magnetically ordered structures, namely, the solitons of the envelope of magnetostatic waves (MSWs) in garnet-ferrite films [2], have been studied extensively. Interest in MSW solitons stems both from the possibility of implementing various integrated spin-wave devices based on them and from the variety of soliton dynamical regimes realizable at fairly low MSW intensities. The generation conditions for solitons of various MSW types and the analytical apparatus for their investigation are described in [2, 3]. Theoretical and experimental studies of MSW solitons in garnet-ferrite films with various orientations of the magnetizing field as well as the types of excitation and control of the soliton dynamics are presented in [4–7]. In particular, it is shown in [7] that the soliton regimes can be controlled through the action of a continuous-pumping wave. Additional possibilities for controlling the soliton regimes could be associated with the use of two-layer magnetically coupled structures as a wave-guiding medium, since the dynamical properties of the spin subsystem change significantly in such structures and new types of spin-wave excitations are realized [8–10]. The dipole [11, 12] or exchange [13] interaction or their combined action [14] are the main types of interlayer magnetic coupling. Recently, the propagation of optical solitons has been investigated in two-channel and two-mode optical fibers [15, 16]. The existence of intermode coupling was shown to lead to energy transfer and oscillations of the amplitudes of the coupled solitons with a certain period along the optical fiber length. The peculiarities of the waveguide structures for MSWs consist mainly in the mode excitation methods and in the constraints imposed on the waveguide length, which, however, is compensated for by the strong nonlinearity of systems. In this paper, we investigate the quasi-solitons of the envelope of forward body MSWs that propagate in a normally magnetized two-layer garnet-ferrite structure under the initial excitation of an MSW soliton in the structure only in one of the coupled magnetic layers.

2. BASIC EQUATIONS

Let us consider a three-layer planar structure composed of two ferromagnetic films separated by a nonmagnetic interlayer that is oriented perpendicular to the z axis. Let a wave packet formed by the interacting MSW modes belonging to each of the films be excited in this film structure and propagate along the x axis. In this case, the magnetostatic potential of the wave packet can be represented as the sum of eigenmodes of the isolated magnetic films in the structure. Of the complete set of film eigenmodes, the only two effectively coupled ones are those that most precisely satisfy the phase matching conditions, i.e., those for which both $2\delta_\omega = \omega_{01} - \omega_{02}$ and $2\delta_k = k_{01} - k_{02}$ have minimal or zero values, where ω_{0n} and k_{0n} are the carrier frequency and wavenumber of an eigenmode. In this paper, we do not assume the presence of a different perturbation (e.g., a periodic nonuniformity of the layer parameters) that could synchronize the modes of different orders. Therefore, the total pulses are formed only by the coupled modes of the same order belonging to different waveguide layers. In particular, the modes of the first order must be these modes, since they are least absorbed. In this case, the magnetostatic potential is described by

$$\Psi = \sum_{n=1,2}\Psi_n = \sum_n \frac{4\pi M_{0n}}{k_{0n}\sqrt{\chi_n^2+\chi_{an}^2}}U_n(z)\varphi_n(x,t) \times \exp[-i(\omega_{0n}t - k_{0n}x - \phi_n)], \qquad (1)$$

where M_{0n} is the magnetization of the corresponding film and ϕ_n is the initial (at $t = 0$) phase of the wave localized in layer n. For the orientation of the magnetizing field $\mathbf{H}$ under consideration ($\mathbf{H} \parallel \mathbf{M}_{0n} \parallel z$), the diagonal and nondiagonal components of the magnetic susceptibility tensor are defined by the expressions

$$\chi_n = \frac{\omega_{Mn}\omega_{Hn}}{\omega_{Hn}^2 - \omega_{0n}^2}, \quad \chi_{an} = \frac{\chi_n\omega_{0n}}{\omega_{Hn}},$$

where

$$\omega_{Mn} = 4\pi\gamma M_{0n}, \quad \omega_{Hn} = \gamma H - \omega_{Mn}.$$

The profile function $U_n(z)$, which defines the distribution of the mode in each of the n layers over its cross-section, and the coefficient in front of it can be determined from the expression that relates the magnetostatic potential to the variable MSW magnetic field and from the Walker equation:

$$\mathrm{grad}\Psi_n = \mathbf{h}_n, \quad \mathrm{div}(\overleftrightarrow{\mu}_n\nabla\Psi_n) = 0, \qquad (2)$$

where $\overleftrightarrow{\mu}_n$ is the magnetic permeability tensor for film n. The dimensionless complex amplitude of the magnetostatic potential can be expressed in terms of the normalized complex values of the variable magnetic moment components,

$$\varphi_n = \frac{\sqrt{m_{xn}^2 + m_{yn}^2}}{M_{0n}},$$

it is a slowly varying function of the coordinate and time due to the intermode coupling. The dispersion relation for the MSW eigenmodes is

$$\tan(k_{zn}l_n) = -\frac{2\sqrt{-\chi_n - 1}}{\chi_n + 2}, \qquad (3)$$

where l_n is the thickness of the corresponding film, and the transverse wave number is related to the propagation constant by $k_{zn} = k_{0n}\sqrt{-\chi_n - 1}$.

The dynamics of the envelope of each of the interacting MSWs can be represented by the following equation (see the Appendix):

$$i\frac{\partial\varphi_n}{\partial t} + iv_n\frac{\partial\varphi_n}{\partial x} + \frac{d_n}{2}\frac{\partial^2\varphi_n}{\partial x^2} = q_n\exp[\pm 2i(\delta_\omega t - \delta_k x - \delta_\phi)]\varphi_{3-n} + g_{sn}|\varphi_n|^2\varphi_n. \qquad (4)$$

Here, the upper and lower signs in the argument of the exponential refer to the modes with the indices $n = 1$ and 2, respectively; $2\delta_\phi = \phi_1 - \phi_2$ is the initial phase difference between the coupled modes; the dispersion parameters $v_n = \partial w_n/\partial k_n$ and $d_n = \partial^2\omega_n/\partial k_n^2$ obtained at $k_n = k_{0n}$ define, respectively, the MSW group velocity and group velocity dispersion, $g_{sn} = \partial\omega_n/\partial|\varphi_n|^2$ is the nonlinear self-action of the system; and q_n is the intermode coupling coefficient. In the absence of losses, the following condition is satisfied:

$$|\varphi_1|^2 + |\varphi_2|^2 = \mathrm{const},$$

which yields the equality $q_1 = q_2^* = q$ for the coupling coefficients. Below, we take into account the phase mismatch attributable only to the difference between the wavevectors of the two modes, $\delta_\omega = 0$ and $\delta_k = \delta$, by assuming the carrier frequencies of the MSW pulses in both magnetic layers to be identical; this is justified by the MSW excitation methods used [3]. When going to the running coordinate

$$\xi = x - \frac{2v_1v_2}{v_1 + v_2}t,$$

the system of dynamical equations takes the form

$$i\frac{\partial\varphi_n}{\partial t} \pm i\tilde{v}_n\frac{\partial\varphi_n}{\partial\xi} + \frac{d_n}{2}\frac{\partial^2\varphi_n}{\partial\xi^2} = q_n\exp\left\{\mp 2i\left[\delta\left(\xi + \frac{2v_1v_2}{v_1 + v_2}t\right) + \delta_\phi\right]\right\}\varphi_{3-n} + g_{sn}|\varphi_n|^2\varphi_n, \qquad (5)$$

where

$$\tilde{v}_n = v_n\frac{v_1 - v_2}{v_1 + v_2}.$$

In the absence of intermode coupling (and when going to the corresponding running coordinate $\xi = x - v_nt$), each of the dynamical equations admits of an exact

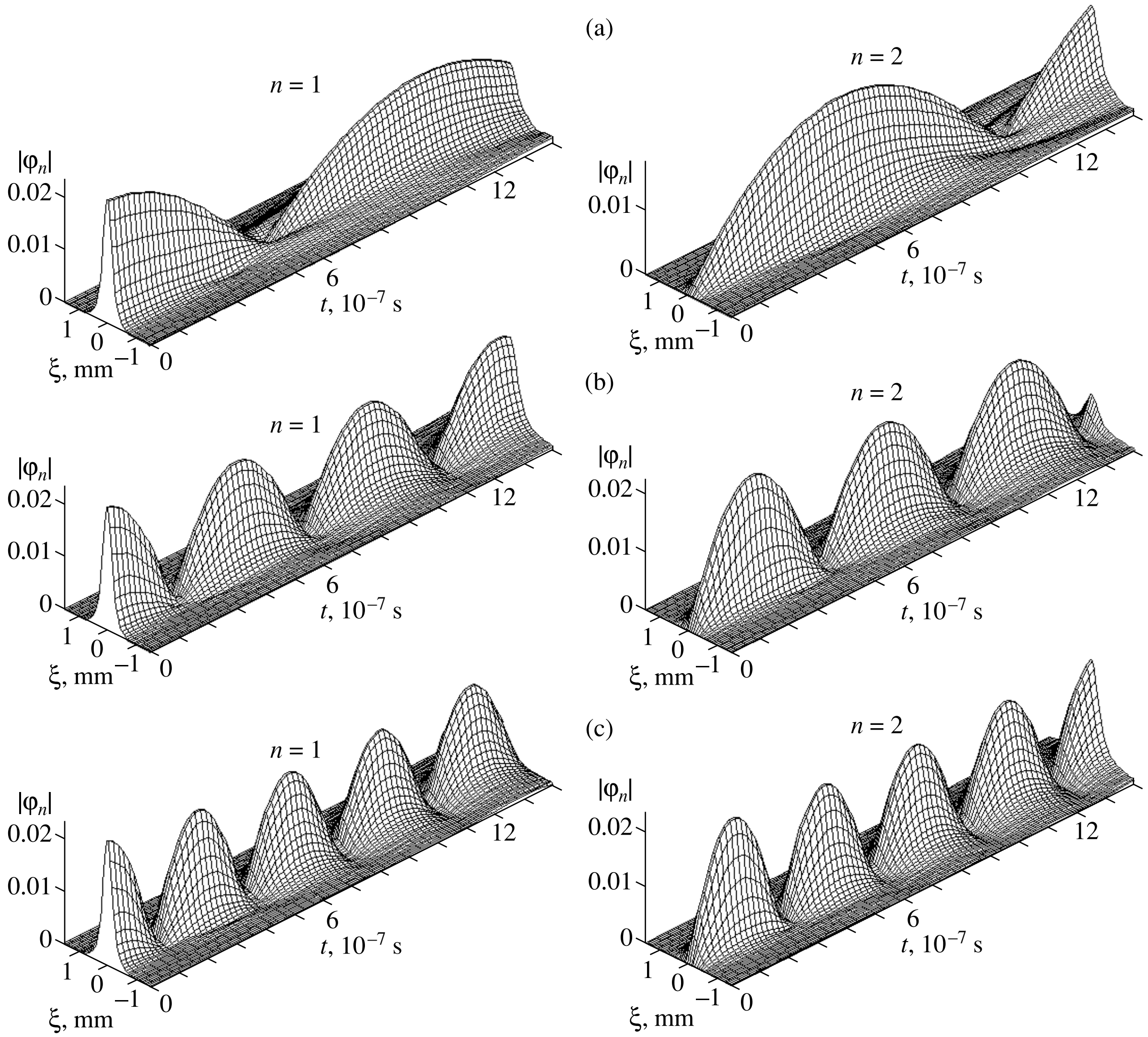


Fig. 1. Time dependences of the envelope of the MSW pulses in each of the layers (n = 1, 2) under the single-mode excitation of the structure and total phase synchronism; $l_{1,2}$ = 5 μm, H = 1.8 kOe, $q = 3 \times 10^6$ (a), 7×10^6 (b), and 10^7 s^{-1} (c).

solution in the form of a propagating soliton of the MSW envelope:

$$\varphi_n(t, \xi) = |\varphi_n(\xi)|\exp(-i\Gamma_n t), \tag{6}$$

where for $d_n < 0$ (anomalous dispersion) and $g_{sn} > 0$,

$$|\varphi_n(\xi)| = |\varphi_{n0}|\cosh^{-1}(\xi/\xi_{n0}). \tag{7}$$

The following relation holds between the initial (i.e., at $t = 0$) amplitude $|\varphi_{n0}|$, length ξ_{n0} of the pulse, and its phase Γ_n:

$$2\Gamma_n = |\varphi_{n0}|^2 g_{sn} = -\frac{d_n}{\xi_{n0}^2}. \tag{8}$$

For normal dispersion ($d_n > 0$) and at $g_{sn} > 0$, we have a stationary solution:

$$\varphi_n(t, \xi) = |\varphi_{n0}|\tanh(\xi/\xi_{n0})\exp(-i\Gamma_n t). \tag{9}$$

In this case, the pulse phase, initial length, and amplitude are related by

$$\Gamma_n = |\varphi_{n0}|^2 g_{sn} = \frac{d_n}{\xi_{n0}^2};$$

this relation describes a state called a "dark" soliton and corresponds to a dip in the MSW intensity.

The solution of Eqs. (5) at $q = 0$ in the form of a propagating soliton is stable [2–4] for the initial amplitude of the solitary wave determined by the dispersion and nonlinear properties of the waveguide and by the length of the pulse excited in the structure. The pulse is

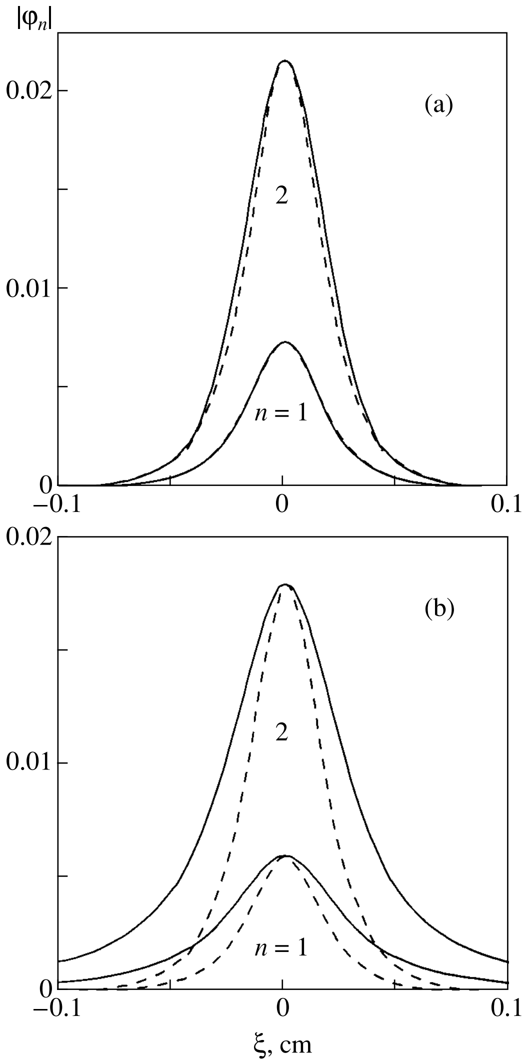


Fig. 2. MSW pulse profiles in each of the layers at (a) $t = 1.26 \times 10^{-7}$ s and (b) $t = 11.34 \times 10^{-7}$ s; the dashed lines represent the soliton profiles with the amplitudes corresponding to the pulses; the coupling coefficient is $q = 10^7$ s^{-1}.

compressed and spreads if the amplitude is, respectively, larger or smaller than $|\varphi_{n0}|$ satisfying condition (8). In actual structures, when the energy of the excited soliton differs slightly from the energy corresponding to the stable solution, the energy is adjusted—it is released in the form of radiation or absorbed from the medium [1].

In the presence of intermode coupling, system (5) can be solved analytically in the case of symmetric or antisymmetric mode excitation in the magnetic layers ($\varphi_{10} = \pm\varphi_{20}$). In the approximation of strong intermode coupling, the envelopes of the totally synchronized interacting MSWs can be represented as a sum of uncoupled partial pulses, and two independent dynamical equations [1, 16] can be solved instead of the system. However, in the case of single-mode excitation of the structure ($\varphi_n \neq 0$, $\varphi_{3-n} = 0$), we cannot obtain two independent equations, and only a numerical solution of system (5) is possible.

3. NUMERICAL ANALYSIS

For our numerical analysis, we take the parameters for each of the magnetic layers of the structures that correspond to garnet-ferrite films: $4\pi M_{0n} = 1750$ G and $\gamma = 1.76 \times 10^7$ (Oe s)$^{-1}$. We assume the duration of the MSW pulse excited in the structure to be $\tau_1 = 10$ ns and the carrier frequency to be $\omega_{0n} = 2 \times 10^9$ s^{-1}.

For these parameters, Fig. 1 shows the time dependences of the MSW pulse profiles in each of the layers of the magnetically coupled structure with a length of $L = 1$ cm for three values of the coupling coefficient and the initial excitation of only the first mode ($n = 1$). The profile of the input pulse corresponds to the soliton solution (7). We also assume that the magnetic layers are identical and have a thickness of $l_n = 5$ μm and that the total phase synchronism condition is satisfied between the MSW modes; the magnetizing field is taken to be $H = 1.8$ kOe. We see from the above dependences that as the input pulse propagates, its energy is transferred to the second layer of the structure, where the pulse of the second MSW mode with a profile similar to the soliton profile is formed. After the pulse of the first mode completely transforms into the pulse of the second mode, the reverse process begins: the amplitude of the $n = 2$ pulse decreases, while the amplitude of the $n = 1$ pulse increases. Depending on the MSW group velocity and the coupling coefficient, this cycle of energy transfer can be repeated many times over the travel time of the pulses through the waveguide, but the pulse profiles slightly change and increasingly deviate from the initial one corresponding to the soliton profile. Thus, two quasi-solitons that "pulsate" as they propagate are formed in the two-layer structure. The number of pulsations of the two synchronized ($v_1 = v_2$) pulses in length L (to within the fractional part) is $N = Lq/\pi v_n$, and the energy from the pulse of one mode is completely transferred to the pulse of the other mode in a time of $t_p = \pi/2q$.

Figure 2 shows the pulse profiles for the two MSW modes at the beginning of their propagation through the waveguide structure at $t = 126$ ns (a), which corresponds to the first half-period of the mode conversion, and at $t = 1134$ ns (b) near the exit from the waveguide for the case considered above at $q = 10^7$ s^{-1}. The dashed lines indicate the profiles that have the same amplitudes as those of the above pulses and that were constructed using formula (7); i.e., they are the soliton profiles. We see that at the beginning of the mode conversion, both the decreasing (in amplitude) pulse and the increasing pulse of the other mode differ little from solitons. However, the mode coupling leads to a spreading of the pulse profiles; as a result, the pulse length at the exit from the waveguide almost doubles compared to the solitons (Fig. 2b). The pulse spreading depends mainly

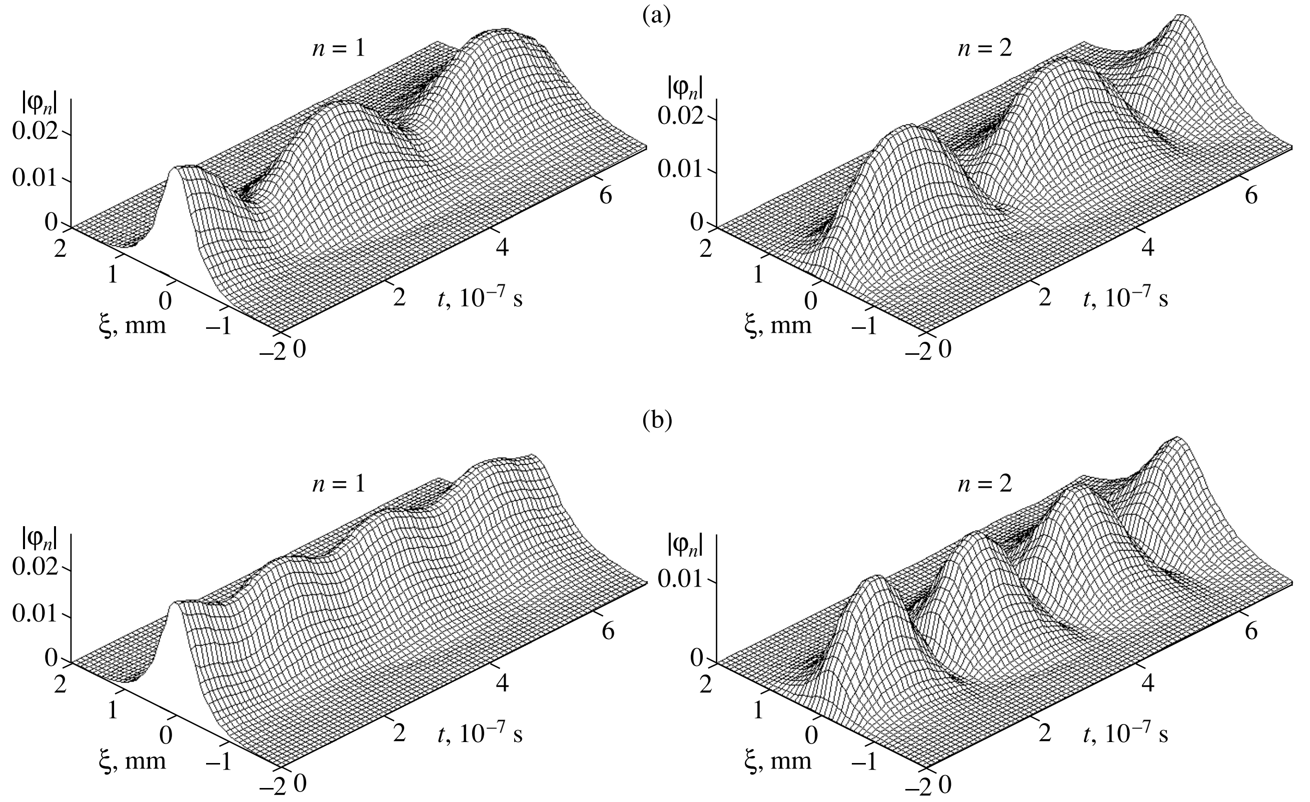


Fig. 3. Time dependences of the envelope of the MSW pulses in each of the layers in the presence of a phase mismatch; l_1 = 10.0 μm, l_2 = 9.9 (a) and 9.7 μm (b), H = 1.8 kOe, and $q = 10^7$ s^{-1}.

on the time of their propagation through the structure. Thus, for example, as the coupling coefficient decreases by several factors, the intermode conversion slows down significantly, but the pulse broadening at the exit from the waveguide changes only slightly. However, as the thickness of the magnetic layers increases, the pulse spreading at the exit from the waveguide proves to be smaller due to an increase in the group velocity of the pulses and, hence, the faster their propagation through the structure. Note that the pulse modulation, which manifests itself in a distortion of the pulse tails, increases with decreasing coupling coefficient (see Fig. 1a).

When performing our numerical analysis, we may not include the initial phase difference δ_ϕ in system (5), since we consider single-mode excitation and the ($n = 2$) mode absent at $t = 0$ lags behind the input ($n = 1$) mode by $\pi/2$ in phase. However, for two-mode excitation, the initial phase difference determines the pattern of intermode coupling: for symmetric or antisymmetric excitation of the structure ($\varphi_{10} = \pm\varphi_{20}$), for which the phase difference is $2\delta_\phi = 0, \pi$, no energy is transferred between the modes, and, hence, there are no periodic mode amplitude variations; at $2\delta_\phi = \pm\pi/2$ for a symmetric waveguide, energy is completely transferred from one mode to the other and back, i.e., the pattern of intermode coupling is similar to that shown in Fig. 1; at intermediate δ_ϕ and at $2\delta_\phi = 0, \pi$ and $\varphi_{10} \neq \pm\varphi_{20}$, mode amplitude pulsations emerge, but there is no complete energy transfer between the modes.

In the case of single-mode excitation under consideration, there is no complete energy transfer between the pulses of the two modes in the presence of a phase mismatch, which, in particular, is attributable to an asymmetry of the waveguide structure. Figure 3 shows the time dependences of the pulse profiles for the two modes at the magnetic layer thickness l_1 = 10.0 μm and l_2 = 9.9 (a) and 9.7 μm (b), the waveguide length L = 1 cm, the coupling coefficient $q = 10^7$ s^{-1}, and the magnetizing field H = 1.8 kOe. As above, the initial excitation is assumed to be single-mode or, more specifically, $|\varphi_2(\xi)| = 0$, and $|\varphi_1(\xi)|$ corresponds to the soliton profile (7). The wavevectors of the MSW eigenmodes in each of the films at the pulse carrier frequency were calculated numerically from the transcendental dispersion relation (3). It follows from these dependences that the mode conversion efficiency decreases with increasing phase mismatch attributable to the difference between the thicknesses of the magnetic layers, which manifests itself in a decrease in the pulsation amplitude of the input pulse and the amplitude of the pulse of the second mode excited by it. In this case, the time of maximum mode conversion decreases; i.e., there is a large

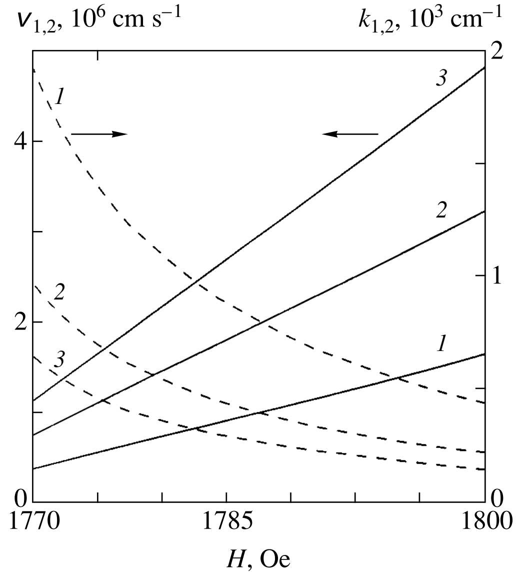


Fig. 4. Magnitudes of the MSW wavevectors (dashed curves) and pulse group velocities (solid curves) versus magnetizing field; $\omega_{0n} = 2 \times 10^9$ s^{-1}, $l_{1,2}$ = 5 (*1*), 10 (*2*), and 15 μm (*3*).

number of extrema of the pulse amplitude over the travel time of the pulses through the waveguide structure.

In Fig. 4, the magnitudes of the MSW wavevectors and the pulse group velocities are plotted against the magnetizing field H at the MSW carrier frequency $\omega_{0n} = 2 \times 10^9$ s^{-1} and the magnetic layer thicknesses l_n = 5, 10, and 15 μm. We see that increasing the magnetizing field causes a decrease in the magnitudes of the wavevectors and a nearly linear increase in the pulse group velocities. As the thickness of the magnetic layers increases, the field dependence of the wavevectors weakens, while the corresponding dependence of the group velocities strengthens. Note that a change in the group velocity at the same pulse duration τ_n causes the pulse length ξ_n to change, but the pulse amplitude for a given change in the magnetizing field changes only slightly due to the accompanying changes in the dispersion and nonlinear self-action of the system (the coefficients d_n and g_{sn}).

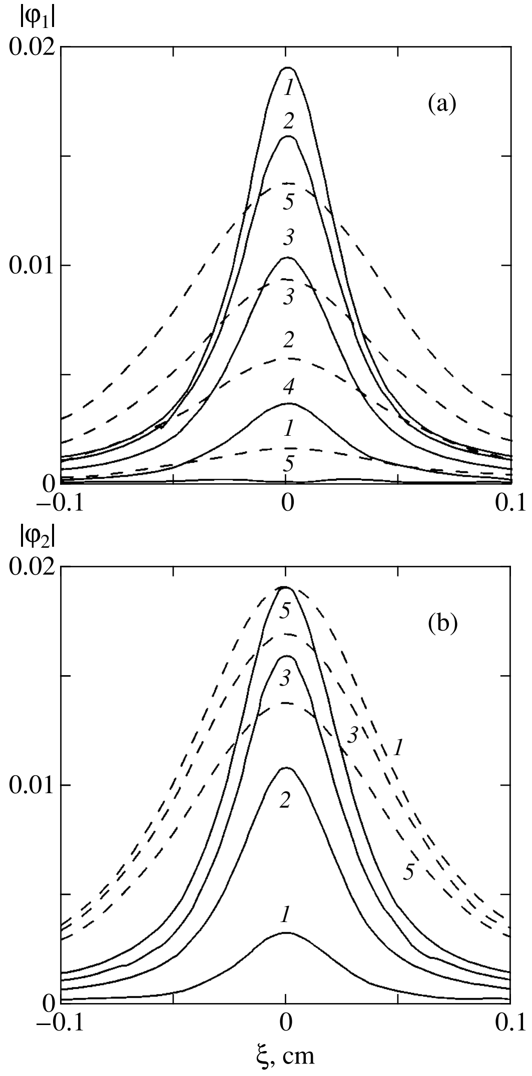


Fig. 5. Pulse profiles in the first (a) and second (b) layers at the exit from a waveguide of length L = 1 cm; $l_{1,2}$ = 6 μm (solid curves) and 10 μm (dashed curves), H = 1796.0 (*1*), 1797.0 (*2*), 1798.0 (*3*), 1799.0 (*4*), and 1799.5 Oe (*5*).

The dependence of the group velocity on the magnetizing field leads to the following important result. By varying the velocity of pulse propagation at the exit from the structure through the field H, we can obtain any mode conversion phase and, thus, regulate the output amplitude of the two MSW pulses over a wide range, from its maximum value to zero in the case of total mode synchronism. Figure 5 shows the profiles of the first (a) and second (b) pulses at the exit from a waveguide of length L = 1 cm for identical magnetic layers of thickness l_n = 5 and 10 μm at various magnetizing fields H. For thicker films, the change in the output amplitudes for the same change in the magnetizing field is smaller. This is because as the films become thicker, the group velocity increases greatly (in the case under consideration, by more than a factor of 2), and the number of extrema of the amplitudes of the propagating pulses decreases over their travel time through the waveguide. As a result, to change the intermode coupling phase by a half-period (which is needed to obtain any of the possible output pulse amplitudes) requires a larger change in the group velocity v_n through the field H than that for thin films; for the latter, the group velocity is relatively low, and the pulse undergoes a much larger number of pulsations as it travels through the waveguide. It is clear from the aforesaid that for a larger change in the field H, several maxima and minima of the pulse amplitude take place at the exit from the magnetic structure; the intervals between the field strengths corresponding to the nearest maxima (minima) of the pulse amplitude in one of the layers increase with increasing H.

4. CONCLUSIONS

Our analysis has shown that when a soliton of the MSW envelope of one of the layers is excited in a two-layer magnetically coupled structure, MSW quasi-solitons pulsating out of phase and belonging to different magnetic layers are formed through intermode coupling. The intermode coupling leads to a spreading of the pulses relative to the soliton profile, which depends mainly on the travel time of the pulses through the two-layer structure. The conversion period of the pulse of one mode into the pulse of the other mode is determined by the coupling coefficient. The dependence of the group velocity on the magnetizing field allows one to obtain any phase of the intermode conversion cycle at the exit from the waveguide structure through a small change in the field strength ($\Delta H \approx 1$ Oe) and, thus, to vary the output amplitude of the pulses of both coupled modes over a wide range. The efficiency of this pulse amplitude control increases with decreasing thickness of the magnetic layers. Note that, despite the MSW peculiarities, the main features of the dynamical behavior of coupled solitons considered here can take place in a broad class of physical systems described by the nonlinear Schrödinger equations.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Education of Russia (project no. NPD02-1.2-72).

APPENDIX

Let us write the wave equation for the magnetostatic potential of an individual layer by assuming it to be isolated (and omitting the index n):

$$\frac{\partial^2\Psi}{\partial x^2} + \frac{\partial^2\Psi}{\partial z^2} - \frac{1}{v^2}\frac{\partial^2\Psi}{\partial t^2} = 0, \tag{A.1}$$

where $v = \bar{\omega}/\bar{k}$ is the MSW phase velocity, $\bar{k} = \sqrt{k^2 + k_z^2}$, and the frequency $\bar{\omega}$ is assumed to depend on the propagation constant k and the MSW intensity. Let us represent the magnetostatic potential as

$$\Psi(t, z, x) = \bar{\Psi}(t, z, x)\exp(ik_0x). \tag{A.2}$$

Its Fourier components,

$$\tilde{\Psi}(t, z, k_0 - k) = \int_{-\infty}^{\infty}\bar{\Psi}(t, z, x)\exp[i(k_0 - k)x]dx, \tag{A.3}$$

satisfy the equation

$$k^2\tilde{\Psi} - \frac{\partial^2\tilde{\Psi}}{\partial z^2} + \frac{\bar{k}^2}{\bar{\omega}^2}\frac{\partial^2\tilde{\Psi}}{dt^2} = 0. \tag{A.4}$$

We seek a solution of this equation in the form

$$\tilde{\Psi}(t, z, k_0 - k) = F(z)\tilde{\varphi}(t, k_0 - k)\exp(-i\omega_0 t), \tag{A.5}$$

where the function $F(z)$ is the product of the profile function $U(z)$ and the coefficient in front of it in Eq. (1); $\tilde{\varphi}(t, k_0 - k)$ are the Fourier components of the MSW envelope $\varphi(t, x)$. Given that

$$\frac{\partial^2 F}{\partial z^2} = (k^2 - \bar{k}^2)F,$$

and discarding the term proportional to $\partial^2\tilde{\varphi}/\partial t^2$, we reduce Eq. (A.4) to the following equation for the slowly varying functions $\tilde{\varphi}$:

$$2i\omega_0\frac{\partial\tilde{\varphi}}{\partial t} - (\bar{\omega}^2 - \omega_0^2)\tilde{\varphi} = 0. \tag{A.6}$$

Let us represent the MSW frequency as

$$\bar{\omega}(k, |\varphi|^2) = \omega(k) + \Delta\omega(|\varphi|^2),$$

where the term $\Delta\omega$ describes the nonlinear effects. Given that the approximate equality $\bar{\omega}^2 - \omega_0^2 \approx 2\omega_0(\bar{\omega} - \omega_0)$ holds good, Eq. (A.6) can be written as

$$\frac{\partial\tilde{\varphi}}{\partial t} = -i[\omega(k) + \Delta\omega - \omega_0]\tilde{\varphi}. \tag{A.7}$$

Let us expand the frequency $\omega(k)$ in a Taylor series near the point k_0:

$$\omega(k) = \omega_0 + (k_0 - k)\omega^{(1)} + \frac{1}{2}(k_0 - k)^2\omega^{(2)} + \ldots, \tag{A.8}$$

where $\omega^{(i)} = [\partial^i\omega/\partial k^i]_{k = k_0}$. The small addition to the frequency associated with nonlinearity can be represented as

$$\Delta\omega = \frac{\partial\omega}{\partial|\varphi|^2}|\varphi|^2 = g_s|\varphi|^2. \tag{A.9}$$

Let us substitute expansion (A.8) in which we discard the terms higher than the second order into Eq. (A.7). Performing the inverse Fourier transform in this equation,

$$\varphi(t, x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\tilde{\varphi}(t, k_0 - k) \times \exp[-i(k_0 - k)x]dk, \tag{A.10}$$

replacing the Fourier transform of $k_0 - k$ with the operator $i(\partial/\partial x)$, and using the expression for $\Delta\omega$, we obtain the nonlinear Schrödinger equation for the MSW envelope of an isolated layer:

$$\frac{\partial\varphi}{\partial t} = \left(-\omega^{(1)}\frac{\partial}{\partial x} + \frac{i}{2}\omega^{(2)}\frac{\partial^2}{\partial x^2} - ig_s|\varphi|^2\right)\varphi. \qquad (A.11)$$

We can formally pass to the equations that include the intermode coupling in the two-layer structure using the matrix equation

$$\frac{\partial}{\partial t}\begin{pmatrix}\Psi_1\\ \Psi_2\end{pmatrix} = \begin{bmatrix}c_{11} & c_{12}\\ c_{21} & c_{22}\end{bmatrix}\begin{pmatrix}\Psi_1\\ \Psi_2\end{pmatrix}, \qquad (A.12)$$

where the unperturbed magnetostatic potential of each of the layers with the initial phase is

$$\begin{aligned}\Psi_n(t, z, x) &= F_n(z)\varphi_n(t, x)\\ &\times \exp[-i(\omega_{0n}t - k_{0n}x - \phi_n)].\end{aligned} \qquad (A.13)$$

In the case of weak linear coupling between the modes where the dispersion relation for the MSW in each of the layers can be assumed to be independent of the wave amplitude in the neighboring layer, the equations for the coupled modes can be written as

$$\begin{aligned}\frac{\partial\varphi_1}{\partial t} &= (c_{11} + i\omega_{01})\varphi_1\\ &+ c_{12}\varphi_2\exp[2i(\delta_\omega t - \delta_k x - \delta_\phi)],\\ \frac{\partial\varphi_2}{\partial t} &= (c_{22} + i\omega_{02})\varphi_2\\ &+ c_{21}\varphi_1\exp[-2i(\delta_\omega t - \delta_k x - \delta_\phi)].\end{aligned} \qquad (A.14)$$

Assuming $c_{nn} + i\omega_{0n}$ to be identical to the operator on the right-hand side of Eq. (A.11), we obtain a system of two coupled nonlinear equations where the coefficients $c_{n,n'}$ define the linear MSW intermode coupling. In the absence of coupling between the modes ($c_{n,n'} = 0$), Eq. (A.11) is valid. For total phase synchronism between the modes and at zero initial phase difference, this system is

$$i\frac{\partial\varphi_n}{\partial t} + iv_n\frac{\partial\varphi_n}{\partial x} + \frac{d_n}{2}\frac{\partial^2\varphi_n}{\partial x^2} = q_n\varphi_{3-n} + g_{sn}|\varphi_n|^2\varphi_n. \qquad (A.15)$$

In the presence of a phase mismatch, we obtain the system of equations (4).

REFERENCES

1. G. Agrawal, *Nonlinear Fiber Optics* (Academic, San Diego, 1995; Mir, Moscow, 1996).
2. A. K. Zvezdin and A. F. Popkov, Zh. Éksp. Teor. Fiz. **84**, 606 (1983) [Sov. Phys. JETP **57**, 350 (1983)].
3. A. V. Vashkovskiĭ, V. S. Stal'makhov, and Yu. P. Sharaevskiĭ, *Magnetostatic Waves in the Super-High Frequency Electronics* (Sarat. Gos. Univ., Saratov, 1993) [in Russian].
4. B. A. Kalinikos, N. G. Kovshikov, and A. N. Slavin, Pis'ma Zh. Éksp. Teor. Fiz. **38**, 343 (1983) [JETP Lett. **38**, 413 (1983)].
5. M. Chen, M. A. Tsankov, J. M. Nash, and C. E. Patton, Phys. Rev. B **49**, 12773 (1994).
6. J. M. Nash, C. E. Patton, and P. Kabos, Phys. Rev. B **51**, 15079 (1995).
7. A. V. Kokin and S. A. Nikitov, Fiz. Tverd. Tela (St. Petersburg) **43**, 851 (2001) [Phys. Solid State **43**, 884 (2001)].
8. R. E. Camley, T. S. Rahman, and D. L. Mils, Phys. Rev. B **27**, 261 (1983).
9. S. L. Vysotskiĭ, G. T. Kazakov, A. V. Maryakhin, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **61**, 673 (1995) [JETP Lett. **61**, 693 (1995)].
10. A. M. Shutyĭ and D. I. Sementsov, Pis'ma Zh. Éksp. Teor. Fiz. **78**, 952 (2003) [JETP Lett. **78**, 480 (2003)].
11. S. L. Vysotskiĭ, G. T. Kazakov, A. V. Maryakhin, and Yu. A. Filimonov, Zh. Tekh. Fiz. **68** (7), 97 (1998) [Tech. Phys. **43**, 834 (1998)].
12. S. V. Tarasenko, Fiz. Tverd. Tela (St. Petersburg) **36**, 2554 (1994) [Phys. Solid State **36**, 1390 (1994)].
13. H. Puszkarski, Surf. Sci. Rep. **20**, 45 (1994).
14. S. L. Vysotskiĭ, G. T. Kazakov, M. L. Kats, and Yu. A. Filimonov, Fiz. Tverd. Tela (St. Petersburg) **35**, 1191 (1993) [Phys. Solid State **35**, 606 (1993)].
15. N. N. Akhmediev and A. Ankevich, *Solitons* (Fizmatlit, Moscow, 2003) [in Russian].
16. I. O. Zolotovskiĭ and D. I. Sementsov, Kvantovaya Élektron. (Moscow) **30**, 794 (2000).

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 100, No. 6, 2005, pp. 1069–1074.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 6, 2005, pp. 1216–1222.
Original Russian Text Copyright © 2005 by Fedoruk.

Two-Photon Nutation Excited in a Two-Level Spin System by Microwave and RF Fields

G. G. Fedoruk

Institute of Physics, University of Szczecin, 70-451 Szczecin, Poland
Sevchenko Institute for Applied Physical Problems, Minsk, 220064 Belarus
e-mail: fedaruk@wmf.univ.szczecin.pl

Received December 21, 2004

Abstract—Two-photon transient nutation is observed in a two-level spin system (E'_1 centers in crystalline quartz) using a transverse microwave field and a linearly polarized rf field oriented along a static magnetic field in the electron paramagnetic resonance. Nutation is excited when the sum of the energies of a microwave photon and a rf photon is equal to the energy difference between two spin states. The two-photon nature of nutation is confirmed by measuring its frequency as a function of the amplitude and frequency of the rf field as well as the amplitude of the microwave field. The amplitude of the effective field of two-photon transitions is measured. It is shown that the decay rate of two-photon nutation is close to the decay rate for one-photon nutation and is determined by the spin–spin interaction between E'_1 centers. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Multiphoton processes in which several photons are emitted or absorbed simultaneously play an important role in nonlinear optics [1]. Although such processes have been known for a long time in magnetic resonance as well, the interest in studying their dynamics (in particular, in two-level systems) has appeared only recently [2–5]. It is well known [3] that real intermediate energy levels are required for multiphoton processes involving a change in the magnetic quantum number $\Delta m_s > 1$ to occur in multilevel spin systems. In two-level spin systems, multiphoton processes with $\Delta m_s = 1$ can take place. Such transitions were studied in a series of recent publications devoted to pulsed NMR and EPR [2–5]. In this case, real intermediate levels do not exist; "dressed" spin states may play the role of such levels [3]. The term "dressed states" is applied to the eigenstates of a spin system in a strong microwave (MW) field having a frequency close to the resonance frequency of this system. The energy levels of a dressed system form a "ladder" of doublets separated by the energy of MW-field photons. At the same time, the energy difference between the states in each doublet is determined by the generalized Rabi frequency.

In the subsequent analysis, we will confine ourselves to multiphoton transitions initiated in two-level spin systems by bichromatic radiation produced by a transverse microwave field and a longitudinal radiofrequency (rf) field [3]. Considering pulsed EPR studies of the dynamics of multiphoton processes induced by the above-mentioned bichromatic radiation, it should be noted that, in particular, two-photon processes of this type can be excited when the MW field is oriented not strictly perpendicularly to the static magnetic field and, hence, an MW field component appears along this field. The experimental setup in which the MW field is tilted at 45° to the static field was repeatedly employed for studying the dynamics of two-photon transitions in two-level systems using nutation and nutation echo [6, 7]. In such experiments, two-photon transitions are excited by two components of the MW field with the same frequency. A similar field configuration has been proposed recently for two-photon excitation in stochastic NMR spectroscopy [2].

When a transverse MW field and a longitudinal rf field are used, the difference in their frequencies is large; this opens new potentialities for two-frequency magnetoresonance spectroscopy and its instrumental implementation. In this type of pulse EPR experiments, which are based on simultaneous absorption or emission of several photons with noticeably different frequencies, two- and three-photon electron spin echo was detected when a nonresonant MW field was used for excitation [4]; in this case, the sum of the energies of a MW photon and one or two rf photons was equal to the resonant frequency of the two-level spin system. Rf-field induced transparency of the matter to MW radiation was observed when a two-level system was excited by a bichromatic field with the frequency of the MW field equal to the resonance frequency of the spin system [5].

An analogous field configuration also made it possible to observe one-photon nutation caused by transitions between the dressed states of a two-level system excited by a resonant MW field [8, 9]. Three-photon nutation between the dressed states of a two-level system has also been studied recently in NMR [4]. In this case, nutation was excited by three circularly polarized photons of a low-frequency rf field, when their frequency was close to one-third of the frequency of one-photon nutation caused by the resonance transverse rf field.

In this study, two-photon nonstationary nutation excited in the EPR by combined action of MW and rf fields is observed for the first time (to our knowledge) in a two-level spin system. In contrast to other coherent nonstationary effects (free induction, echo, etc.), nutation directly reflects the dynamics of quantum transitions during the interaction of radiation with the spin system, while the frequency of nutation provides direct information on the amplitude of the effective field of this interaction. These potentialities of nutation are also illustrated for two-photon transitions in a model two-level spin system (E_1' centers in crystalline quartz).

2. PREDICTIONS OF THE THEORY

Let us consider an electron spin system with spin $S = 1/2$ in a static magnetic field B_0 directed along the z axis in the laboratory system of coordinates and in a linearly polarized microwave field $2B_1\cos(\omega t)$ directed along the x axis and a linearly polarized rf field $2B_2\cos(\omega_{\rm rf}t)$ applied along the z axis. In this case, the Hamiltonian of the system (in units of frequency) has the form

$$H_{\rm lab}(t) = \omega_0 S_z + 2\omega_1\cos(\omega t)S_x + 2\omega_2\cos(\omega_{\rm rf}t)S_z, \quad (1)$$

where $\omega_0 = \gamma B_0$ is the Larmor frequency (resonance frequency of spin transitions), $\gamma = g\beta_e/\hbar$ is the gyromagnetic ratio for the electron, g is the electron g factor, β_e is the Bohr magneton, and $\omega_1 = \gamma B_1$ and $\omega_2 = \gamma B_2$ are the Rabi frequencies for the MW and rf fields, respectively.

For $\omega_0 \gg \omega_1$, only one component (σ^+ photons) of the linearly polarized MW field, which rotates in the direction of the Larmor spin precession, plays a decisive role, while the effect of the second component (σ^- photons) of the MW field, which rotates in the opposite direction, can be ignored.

Under the action of bichromatic radiation considered here, multiphoton transitions of the type $\sigma^+ + k\pi_{\rm rf}$ can be excited; in these transitions, one MW σ^+ photon is absorbed and k rf π photons are absorbed ($k > 0$) or emitted ($k < 0$) simultaneously [5]. Since π photons have zero angular momentum, the number of rf photons participating in such a multiphoton transition is not limited.

For describing multiquantum transitions induced by bichromatic radiation, a transition to a generalized rotating system of coordinates (toggling frame) was used in [5]. In this case, the effective field amplitude calculated to within third-order corrections [5] was found to be

$$\begin{aligned}\omega_{1,k} &= \omega_1\left[J_{-k}(z) + \frac{\omega_1}{2\omega_{\rm rf}}\right.\\ &\quad\left.\times\sum_{l\neq k}\sum_{m\neq 0}\frac{J_{-l}(z)J_{m-l}(z)J_{m-k}(z)}{(l-k)m}\right] \qquad (2)\\ &= \omega_1[c_k^{(1)} + c_k^{(3)}],\end{aligned}$$

where parameter $z = 2\omega_2/w_{\rm rf}$ and $J_n(z)$ is the first-kind Bessel function of order n.

The first coefficient $c_k^{(1)}$ describes the effective field for $\omega_1 \ll \omega_{\rm rf}$. In this case, the effective field amplitude for the resonance MW field ($\omega = \omega_0$) is equal to the sum of the amplitudes of the effective field for all multiphoton processes of the type $-m\pi_{\rm rf} + \sigma^+_{\rm mw} + m\pi_{\rm rf}$:

$$\omega_{1,0} = \omega_1 J_0(z). \qquad (3)$$

In the case of a nonresonant MW field for $\omega_0 = \omega + \omega_{\rm rf}$ ($k = 1$), the amplitude of the effective field for the two-photon transition $\sigma^+ + \pi_{\rm rf}$ and corresponding multiphoton processes of the type $-m\pi_{\rm rf} + \sigma^+_{\rm mw} + (m+1)\pi_{\rm rf}$ is given by

$$\omega_{1,1} = \omega_1 J_{-1}(z). \qquad (4)$$

For $z < 1$ or in the absence of multiphoton processes of the type $-m\pi_{\rm rf} + \sigma^+_{\rm mw} + (m+1)\pi_{\rm rf}$, the effective field of the two-photon transition $\sigma^+ + \pi_{\rm rf}$ can approximately be described by the dependence [3]

$$\omega_{1,1} \approx \frac{\omega_1\omega_2}{\omega_{\rm rf}}. \qquad (5)$$

For large values of ω_1, we must take into account the effect of the third-order term $c_k^{(3)}$.

3. EXPERIMENTAL TECHNIQUE

The field configuration in the laboratory reference frame and the energy level diagram showing the pulsed establishment of two-photon resonance ($\omega_0 = \omega + \omega_{\rm ef}$) between the Larmor frequency and the sum of the frequencies of the MW and rf fields are presented in Fig. 1. We used continuous MW and rf fields; a longitudinal magnetic field pulse established the resonance. The initially equilibrium spin system was in a nonresonant static magnetic field $B = B_0 - \Delta B$. The magnetic field was then changed pulse-wise to the resonant value B_0 at instant $t = 0$. Due to the Zeeman effect, the frequency ω_0' of the quantum transition in the spin system changed as a result of the magnetic field jump $\Delta B =$

$|B - B_0|$ to a value of ω_0 and became equal to the sum of the frequencies of the MW and rf fields (Fig. 1b). During the action of the pulse, the resonance interaction between the total field and the spin system was established. The stabilization time for resonance conditions was about 120 ns and was much shorter than T_2 and $2\pi/\omega_{1,1}$, which ensured the excitation of two-photon nutation.

A continuous rf field and a magnetic field pulse were produced by passing current through the same modulation element in the measuring resonator. In this case, measuring the current producing these fields made it possible to determine the amplitude B_2 of the rf field produced in the region of the sample if the amplitude ΔB of the pulse magnetic field is known. The value of ΔB was determined from the frequency of beats (equal to $\Delta B/\gamma$) in the free induction signal observed after the termination of the magnetic field pulse [10]. This enabled us determine the rf field amplitude B_2 approximately to within 5%.

This technique for forming nutation signals was implemented using the X-band EPR pulse spectrometer specially designed by us [9, 11]. The amplitudes of the MW and rf fields (in frequency units) did not exceed 1.5 MHz. To improve the signal-to-noise ratio, we used multichannel digital summation of signals. The phase of the rf field was not locked with the phase of the MW field. There was no locking of the rf field phase to the beginning of the magnetic field pulse either.

Two-photon nutation was observed for E_1' centers ($S = 1/2$) in crystalline quartz bombarded by neutrons. The EPR spectrum of these centers in the X-band in the direction of the magnetic field parallel to the optical axis of the crystal consists of a single line of width $\Delta B_{pp} = 0.016$ mT. The small width of the EPR line and long relaxation times even at room temperature render E_1' centers in quartz a convenient object for nutation studies [3, 6, 8, 9].

4. RESULTS AND DISCUSSION

The experiments were made at room temperature. The static magnetic field was parallel to the optical axis of the crystal. The duration of negative magnetic field pulses was 10 μs, their amplitude $\Delta B = 0.12$ mT, and the pulse repetition period was 1.25 ms.

Figure 2 shows the dependence of the absorption signal amplitude of one-photon nutation on the magnetic field detuning $\delta/\gamma = (w_0 - \omega)/\gamma$ from the resonance value B_0 during the pulse action.

It is well known that the one-photon nutation signal in the case of a nonuniform magnetic resonance linewidth $\sigma \gg \omega_1$ at $T_2 \ll T_1$ (except for time intervals $t < 1/\omega_1$) can be described by the relation [11, 12]

$$v \propto \omega_1 f(\omega) J_0(\omega_1 t) \exp\left(-\frac{t}{2T_2}\right). \tag{6}$$

Here, $f(\omega)$ is the value of normalized function describing the shape of the line in the vicinity of the center of the nonuniform line, $J_0(\omega_1 t)$ is the zero-order Bessel function, and T_1 and T_2 are the spin–lattice and spin–spin relaxation times, respectively. In the case of E_1' centers, the condition $\omega_1 > \sigma$ was satisfied and the sig-

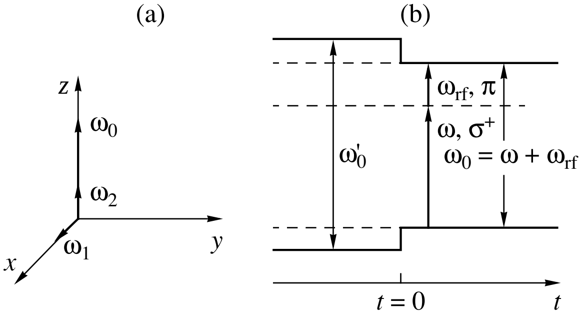


Fig. 1. (A) Field configuration in the laboratory reference frame, used for observing two-photon resonance. (b) Energy level diagram illustrating the pulsed switching on of two-photon resonance ($\omega_0 = \omega + \omega_{rf}$) by a magnetic field pulse in the case of continuous MW and rf fields.

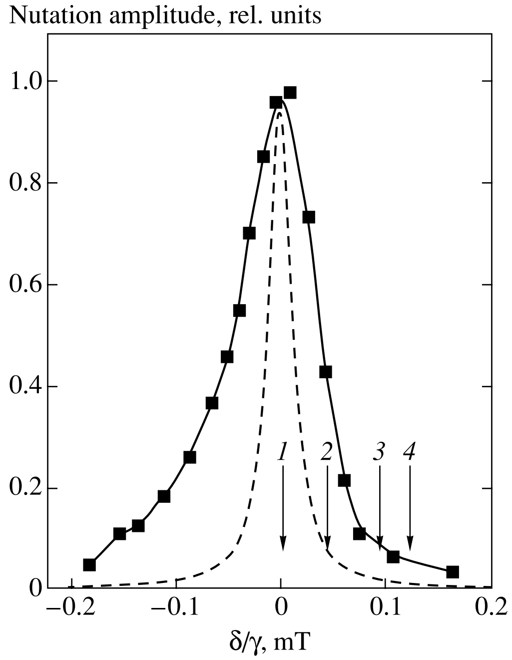


Fig. 2. EPR spectrum of E_1' centers in crystalline quartz, detected with the help of one-photon nutation at $\omega_1/2\pi = 0.96$ MHz. The dashed curve describes the spectrum obtained in a weak microwave field.

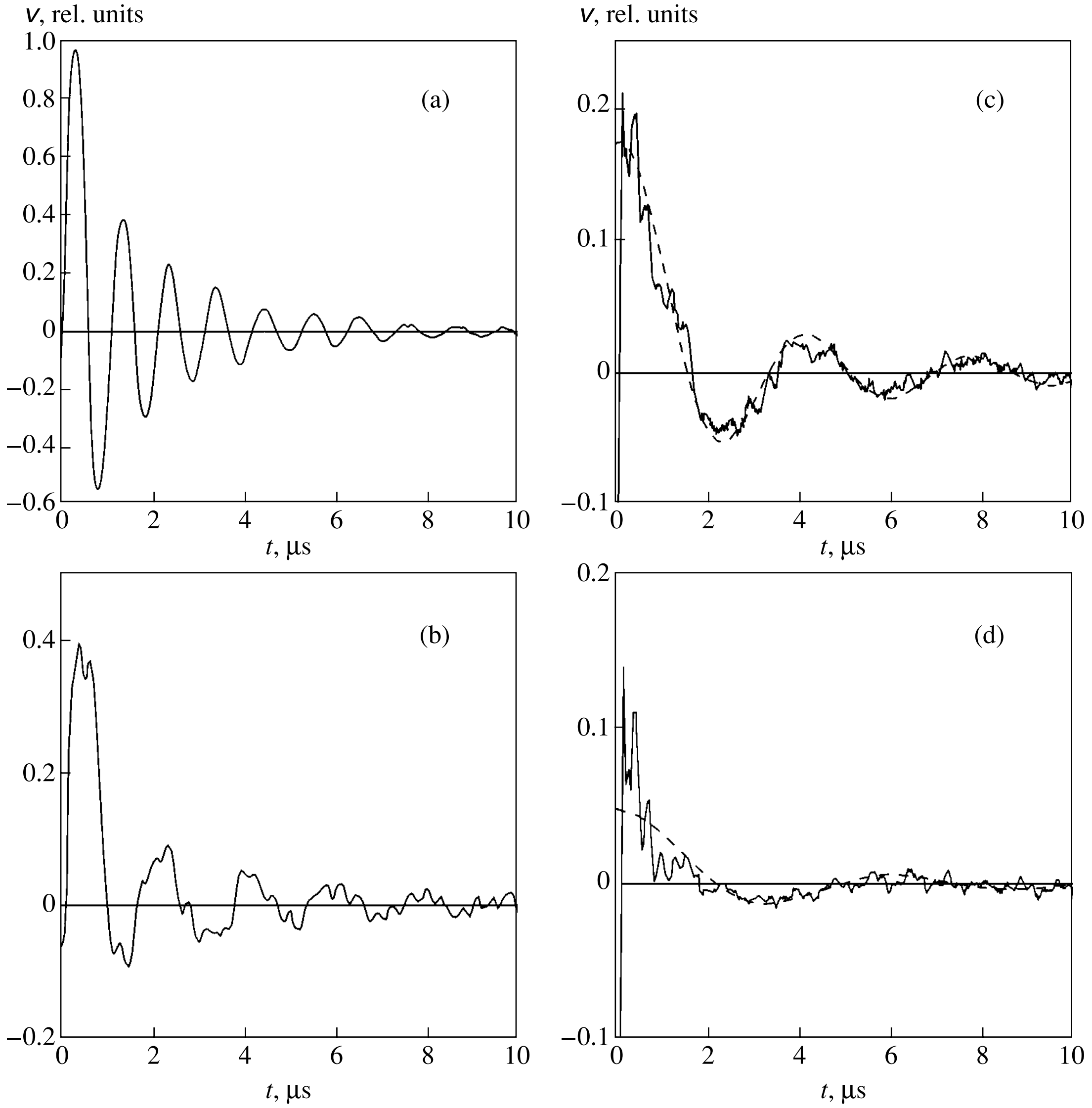


Fig. 3. Nutation EPR signals from E_1' centers detected at fixed amplitudes of the MW ($\omega_1/2\pi = 0.96$ MHz) and rf ($\omega_2/2\pi = 1.02$ MHz) fields for various values of detuning $\delta = \omega_0 - \omega$ from one-photon resonance. (a) One-photon nutation for $\delta = 0$ and $\omega_2 = 0$. Two-photon nutation for detuning $\delta \approx \omega_{rf}$ and $\omega_{rf}/2\pi = 1.37$ (b), 2.62 (c), and 3.40 MHz (d).

nal had a shape intermediate between a decaying Bessel function and a decaying sinusoid typical of a uniformly broadened line (see the oscillograms in Fig. 3a). In this case, the nutation frequency was equal to the Rabi frequency $\omega_1 = \gamma B_1$ and was virtually independent of the detuning from resonance.

In view of radiation broadening, the linewidth detected with the help of nutation exceeds the EPR linewidth for E_1' centers, which can be obtained using a weak MW field and shown by the dashed curve in Fig. 2.

Arrows in Fig. 2 mark the values of the magnetic field during the action of the pulse, at which the nutation signals shown in Fig. 3 were recorded. The signal in Fig. 3a is one-photon nutation at frequency $\omega_1 = 0.96$ MHz, recorded at the resonance value of the magnetic field (the position of this field is marked by arrow *1* in Fig. 2). The signals depicted in Figs. 3b–3d were obtained for magnetic field detunings from the resonance value for one-photon resonance, marked by arrows *2–4* in Fig. 2, for the simultaneous action of the MW and rf fields with frequencies satisfying the condition $\omega + \omega_{rf} = \omega_0$. The rf field amplitude $\omega_2 = \gamma B_2 = 1.02$ MHz was close to the amplitude of the MW field and was the same for all above-mentioned values of detuning from resonance.

It can be seen that, in accordance with relation (5), the frequency Ω_n of nutation observed on the oscillograms shown in Figs. 3b–3d decreases with increasing detuning of the magnetic field and, accordingly, with increasing frequency of the rf field, which is required for the condition $\omega + \omega_{rf} = \omega_0$ to be satisfied. In this case, the amplitude of the recorded signals decreases. It should be noted that if the signal being detected were due not to two-photon transitions, but to a change in the one-photon nutation signal of a uniformly broadened line associated with detuning from resonance, its frequency would be determined by the generalized Rabi frequency $\Omega = \sqrt{\omega_1^2 + \delta^2}$ and would increase with detuning δ.

Figure 4 shows the dependence of the two-photon nutation frequency on parameter $z = 2\omega_2/\omega_{rf}$ (normalized amplitude of the rf field) for $\omega_1/2\pi = 0.96$ MHz. The experimental results represented by circles are obtained upon a change in ω_2 at a fixed rf field frequency $\omega_{rf}//2\pi = 2.62$ MHz. The remaining data are obtained upon a change in the rf field frequency ω_{rf} at $\omega_2/2\pi = 1.02$ MHz. The solid line shows the approximation of the experimental data by the dependence $\Omega_n = k\omega_1\omega_2/\omega_{rf}$, where $k = 0.74 \pm 0.07$.

In view of the absence of phase locking between the rf and MW fields in our experiments, the signal being recorded is the result of averaging of a large number (up to 10^3) of signals obtained for a random phase of the rf field. Such an averaging ensures the suppression of the contribution from all multiphoton processes of the type $-m\pi_{rf} + \sigma^+_{mw} + (m + 1)\pi_{rf}$ [5]. Consequently, the detected nutation was determined by the two-photon transitions $\sigma^+ + \pi_{rf}$ alone, and its frequency can be described by relation (5).

The experimental data presented in Fig. 4 confirm the two-photon origin of the observed nutation. Another dependence confirming this origin is depicted in Fig. 5. It can be seen from Fig. 5 that the two-photon nutation frequency for $\omega_2/2\pi = 1.02$ MHz and $\omega_{rf}/2\pi = 2.62$ MHz is proportional to the MW field amplitude (in frequency units) in accordance with the predictions of the theory.

The possibility of measuring the MW and rf field amplitudes in the given experiment has made it possible to quantitatively compare the effective field of two-photon transitions determined by the nutation frequency with the predictions of the theory (with an error not exceeding 10% even for the lowest nutation frequencies). It has been found that the amplitude of this field differs quantitatively from the value predicted by relation (5) and is close to $k = 1/\sqrt{2} \approx 0.707$. This can be due to the fact that, according to [13, 14], the effect of the longitudinal field on the transitions involving dressed states in such experiments may be weaker than the effect of the transverse field by a factor of $\sqrt{2}$.

It can be seen from Figs. 3c and 3d that the observed signals of two-photon nutation are successfully approximated by a dependence of type (6) typical of a nonuniform line, which can be written in the present case in the form

$$V \propto J_0(\omega_{1,1}t)\exp\left(-\frac{t}{2T_2}\right). \tag{7}$$

The dashed curves in these figures show the depen-

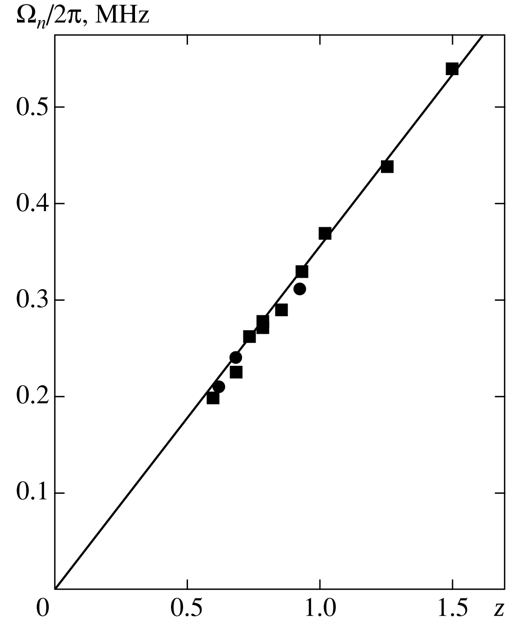


Fig. 4. Dependence of the two-photon nutation frequency on parameter $z = 2\omega_2/\omega_{rf}$.

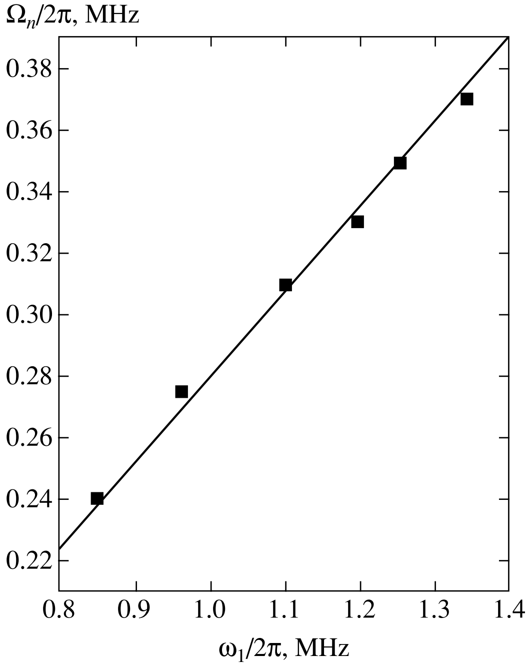


Fig. 5. Dependence of the two-photon nutation frequency on the MW field amplitude (in frequency units) for $\omega_2/2\pi$ = 1.02 MHz and $\omega_{rf}/2\pi$ = 2.62 MHz.

dence given by formula (7) for $\omega_{1,1}/2\pi = 0.28$ MHz (Fig. 3c), $\omega_{1,1}/2\pi = 0.19$ MHz (Fig. 3d) and $T_2 = 3.5$ μs. In this case, the condition $\omega_{1,1} \ll \sigma$ is satisfied more exactly than in the case of one-photon nutation. In view of nonuniform broadening, the two-photon nutation frequency was virtually independent of the detuning from resonance and was determined by the effective field amplitude.

Time $T_2 = 3.5 \pm 0.4$ μs estimated from the two-photon nutation decay coincides with the analogous time for one-photon nutation, which is determined by the dipole–dipole interaction of E'_1 centers [9]. A similar result was observed for E'_1 centers in amorphous quartz in the case of two-photon nutation excited by two MW photons of the same frequency for small amplitudes of the MW field [15]. In this case, the increase in the MW field amplitude led to an anomalous field-dependent decay of two-photon nutation, whose origin is still not completely clear [16]. Probably, two-photon nutation excited by photons with strongly different frequencies will make it possible to obtain new data explaining the reason for the above-mentioned decay.

5. CONCLUSIONS

Thus, we have reported the results of direct recording of two-photon nutation excited in a two-level spin system with a transverse microwave field and a longitudinal radiofrequency field. The possibility of measuring the amplitudes of the MW and rf fields in our experiment enabled us to determine the effective field of two-photon transitions. Apart from the applications illustrated earlier and associated with the transparency induced by the rf field [5] and the replacement of the second MW field in the double electron–electron resonance by a longitudinal rf field [17], analysis of the dynamics of multiphoton transitions using the field configurations employed by us may prove useful for studying relaxation processes in strong exciting fields and for developing new methods for narrowing lines [4].

ACKNOWLEDGMENTS

The author is grateful to I.Z. Rutkovskiĭ for his assistance in experiments.

REFERENCES

1. C. Cohen-Tannoudji, J. Dupont-Roc, and G. Grynberg, *Atom–Photon Interaction: Basic Processes and Application* (Wiley, New York, 1992).
2. C. A. Michal, J. Chem. Phys. **118**, 3451 (2003).
3. I. Gromov and A. Schweiger, J. Magn. Reson. **146**, 110 (2000).
4. H. Hatanaka, M. Sugiyama, and N. Tabuchi, J. Magn. Reson. **165**, 293 (2003).
5. M. Kälin, I. Gromov, and A. Schweiger, Phys. Rev. A **69**, 033809 (2004).
6. R. Boscaino, F. M. Gelardi, and G. Messina, Phys. Rev. A **28**, 495 (1983).
7. G. Bimbo, R. Boscaino, M. Cannas, *et al.*, J. Phys.: Condens. Matter **15**, 4215 (2003).
8. G. Jeschke, Chem. Phys. Lett. **301**, 524 (1999).
9. G. G. Fedoruk, Fiz. Tverd. Tela (St. Petersburg) **46**, 1581 (2004) [Phys. Solid State **46**, 1631 (2004)].
10. V. S. Kuz'min and G. G. Fedoruk, *Nonstationary Coherent Phenomena in Paramagnetic Spin Systems* (Belarus. Gos. Univ., Minsk, 2001) [in Russian].
11. G. G. Fedoruk, Zh. Prikl. Spektrosk. **69** (2), 141 (2002).
12. S. Stoll, G. Jeschke, M. Willer, and A. Schweiger, J. Magn. Reson. **130**, 86 (1998).
13. S. A. Holmstrom, A. S. M. Windsor, C. Wei, *et al.*, J. Lumin. **76**–**77**, 38 (1998).
14. C. Wei, S. A. Holmstrom, A. D. Greentree, and N. B. Manson, J. Opt. B: Quantum Semiclassic. Opt. **1**, 289 (1999).
15. A. Angello, R. Boscaino, M. Cannas, *et al.*, Phys. Rev. A **59**, 4087 (1999).
16. N. Ya. Asadullina, T. Ya. Asadullin, and Ya. Ya. Asadullin, J. Phys.: Condens. Matter **13**, 3475 (2001).
17. M. Fedin, M. Kälin, I. Gromov, and A. Schweiger, J. Chem. Phys. **120**, 1361 (2004).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 100, No. 6, 2005, pp. 1075–1081.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 6, 2005, pp. 1223–1229.
Original English Text Copyright © 2005 by Khriplovich.

NUCLEI, PARTICLES, FIELDS, GRAVITATION, AND ASTROPHYSICS

Radiation of a Quantized Black Hole[¶]

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

Received January 11, 2005

Abstract—The maximum entropy of a quantized surface is demonstrated to be proportional to the surface area in the classical limit. The general structure of the horizon spectrum and the value of the Barbero–Immirzi parameter are found. The discrete spectrum of thermal radiation of a black hole naturally fits the Wien profile. The natural widths of the lines are very small as compared to the distances between them. The total intensity of the thermal radiation is calculated. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The idea of quantizing the horizon area of black holes was put forward many years ago by Bekenstein in the pioneering article [1]. He pointed out that reversible transformations of the horizon area of a nonextremal black hole found by Christodoulou and Ruffini [2, 3] have an adiabatic nature. Of course, the quantization of an adiabatic invariant is perfectly natural, in accordance with the correspondence principle.

Once one accepts this hypothesis, the general structure of the quantization condition for large quantum numbers becomes obvious, up to an overall numerical constant β. The quantization condition for the horizon area A should be

$$A = \beta l_p^2 N, \tag{1}$$

where N is some large quantum number [4]. Indeed, the presence of the Planck length squared,

$$l_p^2 = \frac{k\hbar}{c^3},$$

is only natural in this quantization rule. Then, for the horizon area A to be finite in the classical limit, the power of N must be the same as that of $\hbar$ in l_p^2. This argument can be checked by considering any expectation value in quantum mechanics, nonvanishing in the classical limit. It is worth mentioning that there are no compelling reasons to believe that N is an integer. Neither are there compelling reasons to believe that spectrum (1) is equidistant [5, 6].

On the other hand, formula (1) can be interpreted as follows. The entire horizon area A is split into elements of typical size, $\sim l_p^2$, each of them giving a contribution to the large quantum number N. This scheme arises, in particular, in the framework of loop quantum gravity (LQG) [7–11].

A quantized surface in LQG looks as follows. The surface is assigned a set of edges. Each edge is supplied with an integer or half-integer "angular momentum" j:

$$j = 1/2, 1, 3/2, \ldots. \tag{2}$$

The projections m of these "angular momenta" range as usual from $-j$ to j. The area of the surface is

$$A = 8\pi\gamma l_p^2 \sum_i \sqrt{j_i(j_i+1)}. \tag{3}$$

The numerical factor γ in (3) cannot be determined without an additional physical input. This free (so-called Barbero–Immirzi) parameter [12, 13] corresponds to a family of inequivalent quantum theories, all of them being viable without such an input.

We mention that although spectrum (3) is not equidistant, it is not far from being so. Indeed, even for the smallest quantum number $j = 1/2$, the quantity $\sqrt{j(j+1)}$ can be approximated by $j + 1/2$ with an accuracy of 13%. As j increases,

$$\sqrt{j(j+1)} \approx j + 1/2$$

becomes better and better; i.e., spectrum (3) comes close to being an equidistant one. This feature of spectrum (3) is of interest in connection with the observation by Bekenstein: quantum effects result in the following lower bound on the change of the horizon area ΔA under an adiabatic process:

$$(\Delta A)_{\min} = \xi l_p^2; \tag{4}$$

here, ξ is a numerical factor reflecting "the inherent fuzziness of the uncertainty relation" [14]. Of course,

---

[¶] This article was submitted by the author in English.

the right-hand side of formula (4) is proportional to $\hbar$, together with the Planck length squared l_p^2.

Due to the uncertainty of the numerical factor ξ itself, one cannot see any reason why ξ should not change slightly from one act of capture to another. Therefore, the discussed quasiequidistant spectrum (3) agrees with the bound (4), almost as well as the equidistant one. We return to relation (4) below.

As regards the unknown parameter γ in (3), the first attempts to fix its value, based on an analysis of the black hole entropy, were made in [15, 16]. However, these attempts did not lead to concrete quantitative results.

Then it was argued in [17] that for the black hole horizon, all quantum numbers j are equal to 1/2 (as is the case in the so-called "it from bit" model formulated earlier by Wheeler [18]). With these quantum numbers, one arrives easily at the equidistant area spectrum and at the value

$$\gamma = \frac{\ln 2}{\pi\sqrt{3}}$$

for the Barbero–Immirzi parameter. However, the result in [17] was demonstrated in [5] to be certainly incorrect[1] because it violates the so-called holographic bound formulated in [22–24]. According to this bound, among the spherical surfaces of a given area, the surface of the black hole horizon has the largest entropy.

2. MICROCANONICAL ENTROPY OF A BLACK HOLE

On the other hand, the requirement of maximum entropy allows one to find the correct structure of the horizon area [25], which in particular is of crucial importance for the problem of radiation of a quantized black hole.

We actually consider the "microcanonical" entropy S of a quantized surface defined as the logarithm of the number of states of this surface for a fixed area A (instead of a fixed energy in common problems). Obviously, this number of states K depends on assumptions concerning the distinguishability of edges.

To analyze the problem, it is convenient to rewrite formula (3) as

$$A = 8\pi\gamma l_p^2 \sum_{jm} \sqrt{j(j+1)}\,\nu_{jm}, \tag{5}$$

where ν_{jm} is the number of edges with given j and m. It can be demonstrated [5, 6] that the only reasonable assumption on the distinguishability of edges that may result in acceptable physical predictions (i.e., may comply both with the Bekenstein–Hawking relation and with the holographic bound) is as follows:

$$\text{nonequal } j, \text{ any } m \longrightarrow \text{distinguishable};$$

$$\text{equal } j, \text{ nonequal } m \longrightarrow \text{distinguishable};$$

$$\text{equal } j, \text{ equal } m \longrightarrow \text{indistinguishable}.$$

Under this assumption, the number of states of the horizon surface for a given number ν_{jm} of edges with momenta j and their projections $j_z = m$ is obviously given by

$$K = \nu! \prod_{jm} \frac{1}{\nu_{jm}!}, \tag{6}$$

where

$$\nu = \sum_j \nu_j, \quad \nu_j = \sum_m \nu_{jm},$$

and the corresponding entropy is

$$S = \ln K = \ln(\nu!) - \sum_{jm} \ln(\nu_{jm}!). \tag{7}$$

The structures of the last expression and of formula (5) are so different that the entropy cannot be proportional to the area in the general case. However, this is the case for the maximum entropy in the classical limit.

In this limit, with all the effective "occupation numbers" being large, $\nu_{jm} \gg 1$, we use the Stirling approximation, and hence the entropy is

$$S = \nu \ln \nu - \sum_{jm} \nu_{jm} \ln \nu_{jm}. \tag{8}$$

We calculate its maximum for a fixed area A, i.e., for a fixed sum

$$N = \sum_{jm}^{\infty} \sqrt{j(j+1)}\,\nu_{jm} = \text{const}. \tag{9}$$

The problem reduces to the solution of the system of equations

$$\ln \nu - \ln \nu_{jm} = \mu\sqrt{j(j+1)}, \tag{10}$$

where μ is the Lagrange multiplier for constraint (9). These equations can be rewritten as

$$\nu_{jm} = \nu \exp(-\mu\sqrt{j(j+1)}), \tag{11}$$

---

[1] Later, the result in [17] was also criticized in [19, 20]. Then an error made in [17] was acknowledged [21]. We demonstrate below that the result in [19, 20] is also incorrect.

or

$$\nu_j = (2j+1)\exp(-\mu\sqrt{j(j+1)})\nu. \tag{12}$$

We now sum expressions (12) over j, and with

$$\sum_j \nu_j = \nu,$$

we arrive at the equation for μ:

$$\sum_{j=1/2}^{\infty}(2j+1)\exp(-\mu\sqrt{j(j+1)}) = 1. \tag{13}$$

Its solution is

$$\mu = 1.722. \tag{14}$$

Strictly speaking, the summation in formula (14) extends not to infinity but to some $j_{\max}$. Its value follows from the obvious condition that none of the ν_{jm} should be less than unity. Then, for $\nu \gg 1$, Eq. (11) gives

$$j_{\max} = \frac{\ln\nu}{\mu}. \tag{15}$$

It is well known that the Stirling approximation for $n!$ has reasonably good numerical accuracy even for $n = 1$. Therefore, formula (15) for $j_{\max}$ is not just an estimate but has reasonably good numerical accuracy. The relative magnitude of the corresponding correction to (14) can be easily estimated as $\sim\ln\nu/\nu$.

We now return to Eq. (10). Multiplying it by ν_{jm} and summing over jm, we arrive, with constraint (9), at the following result for the maximum entropy for a given value of N:

$$S_{\max} = 1.722N. \tag{16}$$

Therefore, with the Bekenstein–Hawking relation and formula (5), we find the value of the Barbero–Immirzi parameter:

$$\gamma = 0.274. \tag{17}$$

Quite recently, this calculation with the same result, although with a somewhat different motivation, was reproduced in [26].

We emphasize that the above calculation is not special for LQG only, but applies (with obvious modifications) to a more general class of approaches to the quantization of surfaces. The following assumption is actually necessary here: the surface should consist of sites of different sorts, such that there are ν_i sites of each sort i, with a generalized effective quantum number r_i (here, $\sqrt{j(j+1)}$) and a statistical weight g_i, (here, $2j + 1$). Then in the classical limit, with given functions r_i and g_i, the maximum entropy of a surface can be found, at least numerically, and is certainly proportional to the area of the surface.

As regards the previous attempts to calculate γ, we should point out an apparent error in counting states made in [19, 20]. It can be easily checked that the transition from formula (25) to formulas (29) and (36) in [19] performed therein and then used in [20], is certainly valid under the assumption that only two maximum projections $\pm j$ are allowed for each quantum number j, but it cannot then hold for the correct number $2j + 1$ of the projections. Therefore, it is not surprising that the equation for the Barbero–Immirzi parameter in [20] is

$$2\sum_{j=1/2}^{\infty}\exp(-\mu\sqrt{j(j+1)}) = 1, \tag{18}$$

instead of ours in (13) (see also the discussion of (18) in [26]).

The conclusion is obvious. Any restriction on the number of admissible states for the horizon, as compared to a generic quantized surface, be it the restriction to

$$j = 1/2, \quad m = \pm 1/2,$$

made in [17], or the restriction to

$$\text{any } j, \quad m = \pm j,$$

made in [19, 20], results in a conflict with the holographic bound.

3. QUANTIZATION OF ROTATING BLACK HOLE

In discussing the radiation spectrum of quantized black holes, one should take the angular momentum selection rules into account. Obviously, radiation of any particle with a nonvanishing spin is impossible if both the initial and final states of the black hole are spherically symmetric. Therefore, to find the radiation spectrum, the quantization rule for the mass of a Schwarzschild black hole must be generalized to that of a rotating Kerr black hole.

To derive the quantization rule for a Kerr black hole, we return to the thought experiment analyzed in [2, 3]. Therein, under adiabatic capture of a particle with an angular momentum j, the angular momentum J of a rotating black hole changes by a finite amount j, but the horizon area A does not change. Of course, under some other variation in the parameters, it is the angular momentum J that remains constant. In other words, we have here two independent adiabatic invariants, A and J, for a Kerr black hole with mass M.

Such a situation is quite common in ordinary mechanics. For example, the energy of a particle with mass m bound in the Coulomb field

$$U(r) = -\frac{\alpha}{r}$$

is

$$E = -\frac{m\alpha^2}{2(I_r + I_\phi)^2}, \tag{19}$$

where I_r and I_ϕ are the respective adiabatic invariants for the radial and angular degree of freedom. Of course, the energy E is in a sense also an adiabatic invariant, but it is invariant only under those variations of parameters that preserve both I_r and I_ϕ. In quantum mechanics, formula (19) becomes

$$E = -\frac{m\alpha^2}{2\hbar^2(n_r + 1 + l)^2}, \tag{20}$$

where n_r and l are the radial and orbital quantum numbers, respectively.

This example prompts the solution of the quantization problem for a Kerr black hole. It is conveniently formulated in terms of the so-called irreducible mass M_{ir} of a black hole, related by definition to its horizon radius r_h and area A as

$$r_h = 2kM_{ir}, \quad A = 16\pi k^2 M_{ir}^2. \tag{21}$$

Together with the horizon area A, the irreducible mass is an adiabatic invariant. In accordance with (3) and (9), it is quantized as

$$M_{ir}^2 = \frac{1}{2}m_p^2 N, \tag{22}$$

where

$$m_p^2 = \hbar c/k$$

is the Planck mass squared.

For a Schwarzschild black hole, M_{ir} coincides with its ordinary mass M, but for a Kerr black hole, the situation is more interesting. Here,

$$M^2 = M_{ir}^2 + \frac{J^2}{r_h^2} = M_{ir}^2 + \frac{J^2}{4k^2 M_{ir}^2}, \tag{23}$$

where J is the internal angular momentum of a rotating black hole.

Now, taking (22) into account, we arrive at the following quantization rule for the mass squared M^2 of a rotating black hole:

$$M^2 = \frac{1}{2}m_p^2\left[\gamma N + \frac{J(J+1)}{\gamma N}\right]. \tag{24}$$

Obviously, as long as the black hole is far from an extremal one, i.e., while $\gamma N \gg J$, we can neglect the dependence of M^2 on J, and the angular momentum selection rules have practically no effect on the black hole radiation spectrum.

As regards the mass and irreducible mass of a charged black hole, they are related by

$$M = M_{ir} + \frac{q^2}{2r_h}, \tag{25}$$

where q is the black hole charge. This formula has a simple physical interpretation: the total mass (or total energy) M of a charged black hole consists of its irreducible mass M_{ir} and of the energy $q^2/2r_h$ of its electric field in the outer space $r > r_h$.

With $r_h = 2kM_{ir}$, relation (25) can be rewritten as

$$M^2 = M_{ir}^2 + \frac{q^4}{16k^2 M_{ir}^2} + \frac{q^2}{2k}. \tag{26}$$

Thus, for a charged black hole, M^2 is quantized as

$$M^2 = \frac{1}{2}m_p^2\left[\gamma N + \frac{q^4}{4\gamma N} + q^2\right]. \tag{27}$$

In fact, relations of this type (even in a more general form, for Kerr–Newman black holes, both charged and rotating) were already presented in the pioneering article [1], although with the equidistant quantization rule for M_{ir}^2, i.e., for the horizon area (see also [14]). More recently, the conclusion that the mass of a quantized black hole must be expressed via its quantized area and angular momentum, was made in the approach based on the notion of the so-called isolated horizons [27, 28].

Here, we do not mention the attempts to quantize rotating and charged black holes that resulted in weird quantization rules for $\hat{J}^2$ and $e^2/\hbar c$.

4. RADIATION SPECTRUM OF QUANTIZED BLACK HOLE

It follows from expression (24) that for a rotating black hole, the radiation frequency ω, which coincides with the loss ΔM of the black hole mass, is

$$\omega = \Delta M = T\mu\Delta N + \frac{1}{4kM}\frac{2J+1}{\gamma N}\Delta J, \tag{28}$$

where ΔN and ΔJ are the respective losses of the area quantum number N and the angular momentum J. Here, in line with (24), we used the identity

$$T = \frac{\partial M}{\partial S} = \frac{1}{8\pi k M}\frac{\partial M^2}{\partial M_{ir}^2} \tag{29}$$

for the Hawking temperature T, as well as formula (23).

In the same way, for a charged black hole, with formula (27), we obtain the radiation frequency

$$\omega = \Delta M = T\mu\Delta N + \frac{1}{4kM}\left(2 + \frac{q^2}{\gamma N}\right)q\Delta q, \tag{30}$$

where Δq is the loss of charge.

We are mainly interested in the first temperature terms in (28) and (30), dominating everywhere except the vicinity of the extremal regime, where $J \longrightarrow \gamma N$, or $q^2 \longrightarrow 2\gamma N$, and $T \longrightarrow 0$. The natural assumption is that the thermal radiation occurs when an edge with a given value of j disappears, which means that

$$\Delta N_j = r_j, \quad \omega_j = T\mu r_j. \tag{31}$$

Thus we arrive at the discrete spectrum with a finite number of lines. Their frequencies start at

$$\omega_{\min} = T\mu\sqrt{3}/2$$

and terminate at

$$\omega_{\max} = T\ln\nu.$$

We recall that

$$j \le j_{\max} = \ln\nu/\mu,$$

and hence the number of lines is not very large, $\sim 10^2$, if the black hole mass is comparable to the mass of the Sun. However, because of the exponential decrease of the radiation intensity with ω or j (see below), the existence of $\omega_{\max}$ and a finite number of lines are not that important.

To substantiate the assumption made, we return to the lower bound (4) on the change of the horizon area under an adiabatic capture of a particle. The presence of the gap (4) in this process means that this threshold capture effectively consists in the increase by unity of the occupation number ν_{jm} with the smallest j, equal to 1/2. If the capture were accompanied by a reshuffle of few occupation numbers, the change of the area could easily be made arbitrarily small. For instance, one could delete two edges with quantum numbers j_1 and j_2, and add an edge with the quantum number $j_1 + j_2$. Obviously, with $j_{1,2} \gg 1$, the increase in area could be made arbitrarily small.

It is only natural to assume that in the radiation process as well, changing several occupation numbers instead of one is at least strongly suppressed. We thus arrive at Eqs. (31).

Our next assumption, at least as natural as this one, is that the probability of radiation of a quantum with the frequency ω_j is proportional to the occupation number ν_j. Correspondingly, the radiation intensity I_j at this frequency ω_j is proportional to $\nu_j\omega_j$:

$$I_j \sim \nu_j\omega_j \sim \nu(2j+1)\omega_j\exp(-\omega_j/T). \tag{32}$$

We compare this expression with the intensity of the blackbody radiation in the Wien limit $\omega/T \gg 1$,

$$I(\omega)d\omega = A\frac{\omega^3}{4\pi^2}\exp(-\omega/T)d\omega, \tag{33}$$

where A is the area of a spherical black body. First of all, our relation (32) for I_j directly reproduces the exponential factor of the Wien spectrum. Next, $d\omega$ in (33) goes over into $(1/2)\mu T$ because the limit $\omega/T \gg 1$ corresponds in our problem to $\sqrt{j(j+1)} \gg 1$, i.e., to

$$\sqrt{j(j+1)} \approx j + 1/2,$$

and the minimum increment of j is 1/2. Now, to reproduce the Wien profile, we supplement relation (32) with the following factors: some "oscillator strength" proportional to ω_j, obvious powers of μT, the Newton constant k (necessary to transform ν into A), and obvious numerical ones. We thus arrive at the final formula for the discrete radiation spectrum of a black hole:

$$I_j = AT^4\frac{\mu^4}{8\pi^2}j\left(j+\frac{1}{2}\right)(j+1) \times \exp(-\mu\sqrt{j(j+1)}). \tag{34}$$

Of course, because Wien spectrum (33) corresponds to $j \gg 1$, we cannot guarantee the exact structure of the j-dependence in formula (34), especially in the preexponential factor. For instance, it would perhaps be equally legitimate to write there

$$j^{3/2}(j+1)^{3/2}$$

instead of

$$j(j+1/2)(j+1).$$

However, this ambiguity is not very significant, at least numerically.

We note that because the black hole temperature T is less than the minimum allowed frequency $\omega_{\min}$, this spectrum has no Rayleigh–Jeans region at all.

Now, the emission probability for a quantum of frequency $\omega_j = T\mu r_j$, i.e., the width of the corresponding line, is

$$\Gamma_j = \frac{I_j}{\omega_j} = AT^3\frac{\mu^3}{8\pi^2}(j+1/2)\sqrt{j(j+1)} \times \exp(-\mu\sqrt{j(j+1)}). \quad (35)$$

The ratio of this natural line width to the distance

$$\Delta\omega_j = \omega_{j+1} - \omega_j \approx \frac{1}{2}\mu T$$

between the lines is very small numerically:

$$\frac{\Gamma_j}{\Delta\omega_j} \approx \frac{\mu^2}{16\pi^3}(j+1/2)\sqrt{j(j+1)} \times \exp(-\mu\sqrt{j(j+1)}) \lesssim 10^{-3}. \quad (36)$$

Thus, the radiation spectrum of an isolated black hole is actually discrete.

Finally, the total radiation intensity of a black hole is

$$I = \sum_j I_j = 0.150AT^4. \quad (37)$$

The numerical coefficient in this expression is close to that in the total intensity of the common thermal radiation, i.e., to the Stefan–Boltzmann constant

$$\pi^2/60 = 0.164.$$

The point is that the Rayleigh–Jeans contribution to the total intensity, which is completely absent in the present spectrum, would be small anyway.

Formulas (34) and (37) describe not only the thermal radiation of bosons, photons, and gravitons, but also the thermal radiation of fermions, i.e., massless neutrinos. However, in the latter case, proper account of the number of polarization states is necessary: for a two-component Dirac neutrino, the numerical factors in formulas (34) and (37) are twice as small.

In fact, it was argued long ago [29] that the discrete thermal radiation spectrum of a black hole, with the equidistant quantization rule for the horizon area, should fit the Wien profile.

On the other hand, our conclusion on the discrete radiation spectrum of a black hole in LQG differs drastically from that of [30], according to which the black hole spectrum in LQG is dense.

As regards the nonthermal radiation of extremal black holes, described by the terms with ΔJ and Δq in Eqs. (28) and (30), these effects are due to tunneling (see a relatively recent discussion of the subject and a detailed list of relevant references in [31, 32]). The loss of charge by a charged black hole is in fact caused by the Coulomb repulsion between the black hole and the emitted particles with the same sign of charge. For a rotating black hole, the reason is the interaction of angular momenta: particles (mainly massless) whose total angular momentum is parallel to that of the black hole are repelled from it.

ACKNOWLEDGMENTS

I appreciate numerous useful discussions with O.P. Sushkov. I am also grateful to J. Bekenstein for correspondence; in particular, he directed my attention to limit (4). A substantial part of this work was done during my visit to the School of Physics, University of New South Wales, Sydney, Australia; I wish to thank UNSW for kind hospitality. The investigation was supported in part by the Russian Foundation for Basic Research (project no. 05-02-16627).

REFERENCES

1. J. D. Bekenstein, Lett. Nuovo Cimento **11**, 467 (1974).
2. D. Christodoulu, Phys. Rev. Lett. **25**, 1596 (1970).
3. D. Christodoulu and R. Ruffini, Phys. Rev. D **4**, 3552 (1971).
4. I. B. Khriplovich, Phys. Lett. B **431**, 19 (1998); gr-qc/9804004.
5. I. B. Khriplovich, Zh. Éksp. Teor. Fiz. **126**, 527 (2004) [JETP **99**, 460 (2004)]; gr-qc/0404083.
6. I. B. Khriplovich, gr-qc/0409031.
7. C. Rovelli and L. Smolin, Nucl. Phys. B **442**, 593 (1995); **456**, 753(E) (1995); gr-qc/9411005.
8. A. Ashtekar and J. Lewandowski, Class. Quantum Grav. **14**, 55 (1997); gr-qc/9602046.
9. R. Loll, Phys. Rev. Lett. **75**, 3048 (1995); gr-qc/9506014; Nucl. Phys. B **460**, 143 (1996); gr-qc/9511030.
10. R. De Pietri and C. Rovelli, Phys. Rev. D **54**, 2664 (1996); gr-qc/9602023.
11. S. Frittelli, L. Lehner, and C. Rovelli, Class. Quantum Grav. **13**, 2921 (1996); gr-qc/9608043.
12. G. Immirzi, Class. Quantum Grav. **14**, L177 (1997); gr-qc/9701052.
13. C. Rovelli and T. Thiemann, Phys. Rev. D **57**, 1009 (1998); gr-qc/9705059.
14. J. D. Bekenstein, in *Cosmology and Gravitation*, Ed. by M. Novello (Atlantisciences, France, 2000); gr-qc/9808028.
15. C. Rovelli, Phys. Rev. Lett. **77**, 3288 (1996); gr-qc/9603063.
16. K. V. Krasnov, Phys. Rev. D **55**, 3505 (1997); gr-qc/9603025; Gen. Relativ. Gravit. **30**, 53 (1998); gr-qc/9605047; Class. Quantum Grav. **16**, 563 (1999); gr-qc/9710006.
17. A. Ashtekar, J. Baez, A. Corichi, and K. Krasnov, Phys. Rev. Lett. **80**, 904 (1998); gr-qc/9710007.

18. J. A. Wheeler, in *Sakharov Memorial Lectures in Physics: Proceedings of the First International Sakharov Conference on Physics*, Ed. by L. V. Keldysh and V. Ya. Feinberg (Moscow, 1991), Vol. 2, p. 751.
19. M. Domagala and J. Lewandowski, Class. Quantum Grav. **21**, 5233 (2004); gr-qc/0407051.
20. K. Meissner, Class. Quantum Grav. **21**, 5245 (2004); gr-qc/0407052.
21. A. Ashtekar, gr-qc/0410054.
22. J. D. Bekenstein, Phys. Rev. D **23**, 287 (1981).
23. G. ’t Hooft, in *Salam Festschrift* (World Sci., Singapore, 1993); gr-qc/9310026.
24. L. Susskind, J. Math. Phys. **36**, 6377 (1995); gr-qc/9710007.
25. R. V. Korkin and I. B. Khriplovich, Zh. Éksp. Teor. Fiz. **122**, 5 (2002) [JETP **95**, 1 (2002)]; gr-qc/0112074.
26. A. Ghosh and P. Mitra, gr-qc/0411035.
27. A. Ashtekar, S. Fairhurst, and B. Krishnan, Phys. Rev. D **62**, 104025 (2000); gr-qc/0005083.
28. A. Ashtekar, C. Beetle, and J. Lewandowski, Class. Quantum Grav. **19**, 1195 (2002); gr-qc/0111067.
29. J. D. Bekenstein and V. F. Mukhanov, Phys. Lett. B **360**, 7 (1995); gr-qc/9505012.
30. M. Barreira, M. Carfora, and C. Rovelli, Gen. Relativ. Gravit. **28**, 1293 (1996); gr-qc/9603064.
31. I. B. Khriplovich, Zh. Éksp. Teor. Fiz. **115**, 1539 (1999) [JETP **88**, 845 (1999)]; gr-qc/9812060.
32. R. V. Korkin and I. B. Khriplovich, Zh. Éksp. Teor. Fiz. **121**, 531 (2002) [JETP **94**, 453 (2002)]; gr-qc/0107101.

Journal of Experimental and Theoretical Physics, Vol. 100, No. 6, 2005, pp. 1082–1093.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 6, 2005, pp. 1230–1243.
Original Russian Text Copyright © 2005 by Mitsen, Ivanenko.

ORDER, DISORDER, AND PHASE TRANSITIONS IN CONDENSED SYSTEMS

Charge Ordering, Superconductivity, and Stripes in Doped La_2CuO_4

K. V. Mitsen and O. M. Ivanenko
Lebedev Physical Institute, Russian Academy of Sciences, Leninskiĭ pr. 53, Moscow, 119991 Russia
e-mail: mitsen@sci.lebedev.ru

Received April 15, 2004

Abstract—The special features of the phase diagrams of $La_{2-x}Sr_xCuO_4$ are considered in terms of the high-temperature superconductivity model according to which the mechanism responsible for the anomalous properties of these compounds is the interaction of electrons with diatomic negative U-centers. A microstructural model that assumes the coexistence of domains with different types of dopant ion ordering is suggested for $La_{2-x}Sr_xCuO_4$. According to this model, the main characteristics of the experimental phase diagrams of $La_{2-x}Sr_xCuO_4$ only reflect square lattice geometric relations and competition between different dopant ordering types. Close agreement between the calculated and experimental "superconducting" and "magnetic" phase diagrams is an important argument in favor of the suggested high-temperature superconductivity model. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

In [1], we suggested a model of high-T_c superconductors based on the suggestion that the mechanism responsible for many of the anomalous properties of these compounds, including high-temperature superconductivity proper, is the interaction of electrons with diatomic negative U-centers (NUCs). We showed that their formation was possible because of the stiff localization of doping charges in the vicinity of impurity ions. Electron pairing responsible for superconductivity appears in this model as a result of the strong renormalization of effective electron–electron interaction when scattering with intermediate virtual NUC bound states is taken into account. This mechanism was suggested for the first time in [2] and repeatedly discussed later on as applied to various systems, including high-T_c superconductors [3–10].

The suggested mechanism of the interaction of electrons with pair states can only be effective if percolation clusters are formed from NUCs. However, generally, at a random distribution of dopant ions, the formation of such extensive clusters is difficult to expect. Our goal is to show that there is peculiar ordering of dopant ions in high-T_c superconductors. This creates conditions for the formation of NUC percolation clusters over a broad concentration range. For this purpose, we suggest a model of ordering that allows the phase diagram of $La_{2-x}Sr_xCuO_4$ to be explained. We also show that, in certain dopant concentration ranges, such ordering in $La_{2-x}Sr_xCuO_4$ is accompanied by the formation of magnetic spin textures that simulate stripe structures [11].

2. THE FORMATION OF NEGATIVE U-CENTERS IN HIGH-T_c SUPERCONDUCTORS

In [1], we suggested a mechanism for the formation of diatomic NUCs on a pair of neighboring copper cations in the CuO_2 plane. This mechanism is essentially as follows. The electronic structure of the insulator phase of high-T_c superconductors in the vicinity of the Fermi energy E_F is known to be well described by the model of a insulator with a gap related to charge transfer. The electron energy spectrum of an undoped high-T_c superconductor is shown in Fig. 1a. In this scheme, the excitation energy Δ_{ct} (about 2 eV) corresponds to electron transfer from oxygen to a neighboring copper ion. The hole is then distributed over four surrounding oxygen ions because their orbitals overlap (Fig. 1b). This excitation (an electron on copper and a hole on the surrounding oxygen ions) resembles the hydrogen ion. Proceeding further, we can suggest that the energy of two excitations must be lower if two such pseudoatoms neighbor one another and as though form a hydrogen pseudomolecule (Fig. 1d). This is possible under certain conditions thanks to the formation of a bound state (of the Heitler–London type) of two electrons on neighboring copper ions and two holes that appear in the nearest environment of this pair of cations.

Apart from purely outward similarity, an additional argument in favor of our model is the observation that the distance between copper cations in high-T_c superconductors is a = 3.7–4.0 Å, which is approximately equal to the product $R_0\varepsilon_\infty$, where R_0 = 0.8 Å is the distance between nuclei in the H_2 molecule and ε_∞ is the high-frequency permittivity equal to 4.5–5.0 for all high-T_c superconductors; that is, naturally created con-

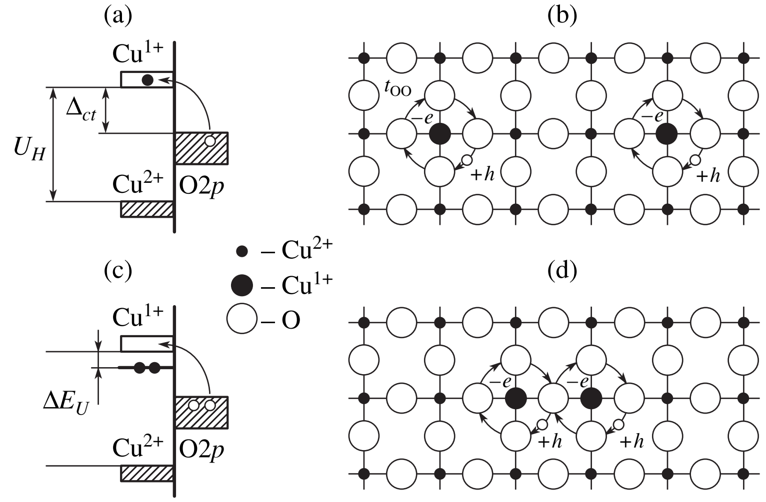


Fig. 1. (a) Electron energy spectrum of an undoped CuO_2 plane: U_H is the energy of repulsion between two electrons on a copper ion; the Δ_{ct} gap for excitation with the lowest energy corresponds to (b) electron transfer from oxygen to the nearest copper ion with the formation of a hole distributed over four surrounding oxygen ions; t_{OO} is the hopping integral between the $p\sigma$ orbitals of the nearest oxygen ions; (c) the energy of two such pseudoatomic excitations can be decreased by ΔE_U if (d) they are situated side by side and form a pseudomolecule.

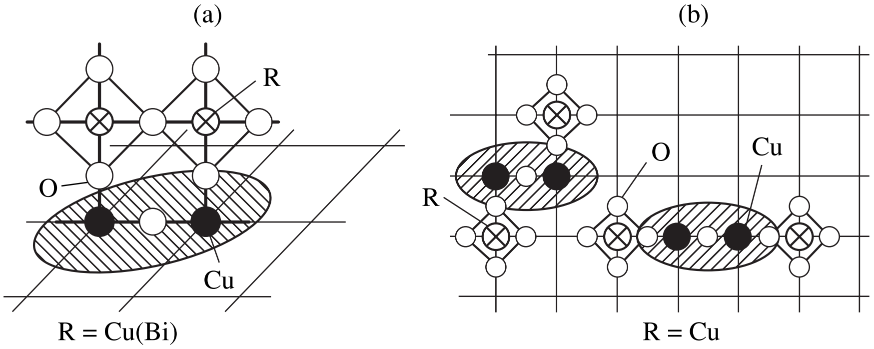


Fig. 2. Atomic clusters $Cu_2R_2O_n$ common to all cuprate high-T_c superconductors with hole doping: (a) in YBCO (BSCCO), copper ions are incorporated into the CuO_2 plane, and R is the Cu (Bi) ion in CuO_3 chains (BiO planes); (b) in $La_{2-x}M_xCuO_4$ (M = Ba or Sr), the whole cluster is built into the CuO_2 plane, and R stands for Cu ions inside the oxygen octahedron adjacent to the M ion. Two types of such clusters with a distance between the R ions of $3a$ or $a\sqrt{5}$ can exist in $La_{2-x}M_xCuO_4$. The ellipses bound the regions of the localization of additional hole pairs that appear when an electron pair is transferred to a NUC.

ditions for the formation of such pseudomolecules in high-T_c superconductor crystals.

Let us consider the conditions under which a NUC is formed on the given pair of copper ions in the CuO_2 plane. For this purpose, we select a fragment of the crystal structure (Fig. 2) common to all cuprates with hole doping. This is the $Cu_2R_2O_n$ cluster, where the copper ions are "built" into the CuO_2 plane and R = Cu in the CuO_2 plane in La_2CuO_4, R = Cu in chains in $YBa_2Cu_3O_{6+\delta}$, and R = Bi in $Bi_2SrCaCu_2O_x$. We assume that a NUC is formed in the CuO_2 plane on a pair of copper ions if a hole formed as a result of doping (doped hole) is localized in each of the oxygen squares surrounding R ions (Fig. 2). For $Y_2Ba_2Cu_3O_{6+x}$, this requirement means that a NUC on a given pair of Cu ions is formed when three consecutive oxygen sites are filled in the CuO_3 chain over these ions. For $La_{2-x}Sr_xCuO_4$, where localized doped holes lie in the CuO_2 plane (on four oxygen atoms of the oxygen octahedron adjacent to the strontium ion [12, 13]), this requirement is fulfilled when the distance between the R ions (or, equivalently, between the projections of strontium ions onto the CuO_2 plane) is $3a$ or $a\sqrt{5}$ (a is the lattice constant in the CuO_2 plane).

In [1], we characterized various types of the mutual arrangement of doped charges in $La_{2-x}Sr_xCuO_4$. According to simple estimates [1], a doped hole localized in the CuO_2 plane decreases the Δ_{ct} value by

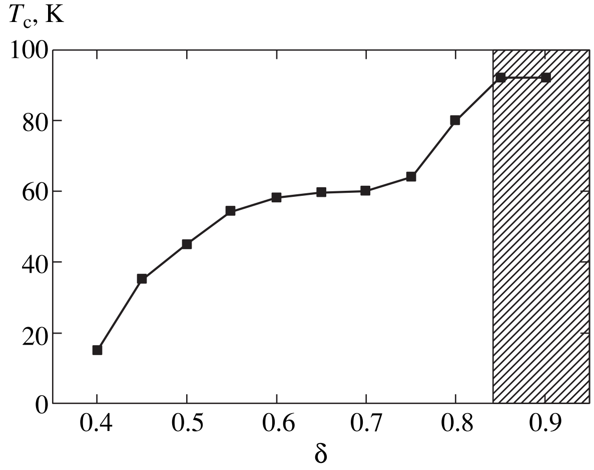


Fig. 3. T_c–δ phase diagram for $YBa_2Cu_3O_{6+\delta}$ [15]. The region of NUC percolation is hatched.

approximately 1.8 eV for the nearest four copper ions. When doped holes localized in the CuO_2 plane inside oxygen octahedra are at a distance of $a\sqrt{5}$ and 3*a* from each other, they decrease the Δ_{ct} value for the internal pair of neighboring copper ions to the extent that two-electron transfer to this pair from the surrounding oxygen ions becomes possible. These two variants of the mutual arrangement of doped charges correspond to the formation of a NUC on a pair of neighboring copper ions.

In the intermediate case, when the distance between the doped charges is $a\sqrt{8}$, there is no pair of neighboring copper ions with a doped hole adjacent to each of them, and NUCs are not formed. This situation corresponds to an insulator and, as is shown below, is responsible for the "1/8"-anomaly.

When doped holes are situated at a distance of 2*a* from each other, the Δ_{ct} gap for the internal copper ion closes also for one-electron transitions. This corresponds to the conventional metal state.

As is seen from Fig. 1, NUCs play the role of pair acceptors that generate additional hole pairs, which are also localized in the neighborhood of NUCs. There is conductivity in such a system if these regions of the localization of hole pairs form a percolation cluster in the CuO_2 plane; that is, when the percolation threshold is exceeded along the NUC chain.

At a random dopant distribution in a crystal, various types of the mutual arrangement of doped charges can coexist. It would then be difficult to expect the existence of large clusters whose properties are fully determined by one of the four configuration variants considered above. We will, however, show that the dopants in $La_{2-x}Sr_xCuO_4$ are ordered over square lattice sites with different constants that depend on their concentration. For this reason, extensive percolation nets of broken lines with segment lengths $a\sqrt{5}$, 3*a*, $a\sqrt{8}$, and 2*a* can develop on such lattices. The first two segment lengths correspond to the conducting (high-T_c) phase in our model; the third one, to an insulator; and the fourth, to a conventional metal.

3. THE CHARGE ORDERING MODEL

3.1. $YBa_2Cu_3O_{6+\delta}$

The ordering of doping oxygen ions is inherent in $YBa_2Cu_3O_{6+\delta}$ because they have crystallographic sites of their own. According to [1], a NUC in a certain $YBa_2Cu_3O_{6+\delta}$ cell is formed when three consecutive oxygen sites are filled in the CuO_3 chain (see Fig. 2). The concentration of such cells is δ^3 if the oxygen ions are distributed at random. Percolation occurs when the site percolation threshold p_c is exceeded for the square lattice; this threshold is 0.593 [14]. Suppose that such oxygen triples are distributed independently (strictly, this is not so). We then have $\delta_c^3 = 0.593$ and $\delta_c = 0.84$, where δ_c is the δ value corresponding to the percolation threshold. In reality, such triples of oxygen ions are distributed not independently, and δ_c therefore differs from 0.84 but insignificantly.

The T_c–δ phase diagram of $YBa_2Cu_3O_{6+\delta}$ samples with different oxygen contents is shown in Fig. 3 [15]. The region $0.84 < \delta < 1$, where percolation along NUC chains occurs, is hatched. We see that the concentration interval corresponding to percolation along NUC chains coincides with the region of high-temperature superconductivity, $T_c \approx 92$ K.

3.2. $La_{2-x}Sr_xCuO_4$

In this section, we suggest and substantiate the model of an ordered distribution of strontium ions over lanthanum sites in La_2CuO_4. This model allows us to explain the T_c–*x* diagram for $La_{2-x}Sr_xCuO_4$ and consider the characteristics of magnetic spin textures formed as a result of ordering.

First, let us consider the special features of the crystal structure of La_2CuO_4 and find out how and where doped charges (holes) are formed. The unit cell of La_2CuO_4 containing two formula units is shown in Fig. 4a. As is seen from this figure, there are two different lanthanum sites in the La_2CuO_4 lattice (for instance, in the LaO planes I and II) shifted by $a/\sqrt{2}$ with respect to each other in the direction of the cell diagonal. It follows from the experimental data [12, 13] that doped holes are rigidly localized in the CuO_2 plane on four oxygen atoms of the oxygen octahedron adjacent to the dopant ion (Fig. 4a); that is, holes are doped into the central CuO_2 plane by ions situated in planes I and IV and are localized in the immediate vicinity of the

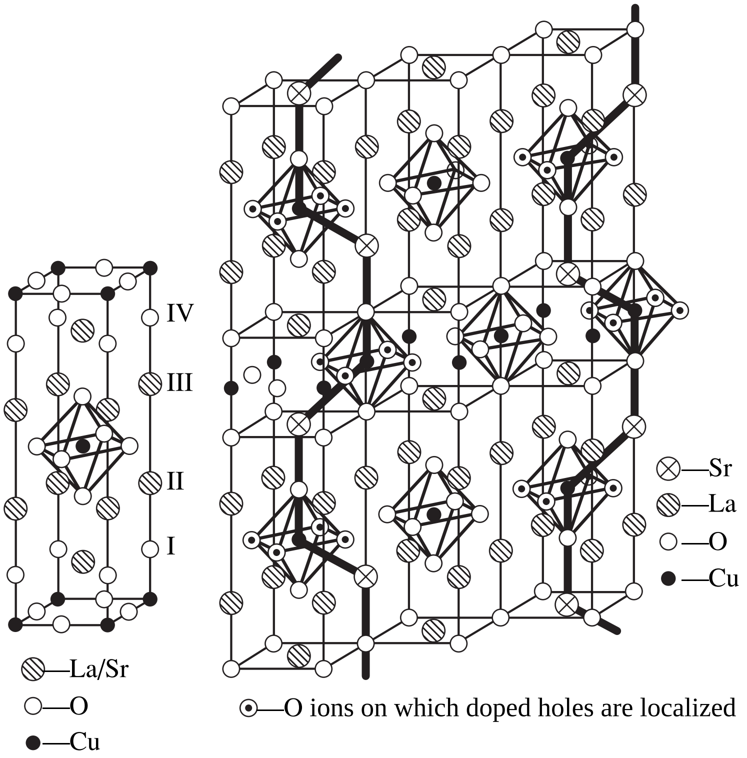


Fig. 4. Ordering of strontium ions in $La_{2-x}Sr_xCuO_4$: (a) La_2CuO_4 unit cell; Roman numerals on the right are the numbers of LaO planes (see text); (b) negatively charged strontium ions together with doped holes "ascribed" to them are dipoles that attract each other at opposite ends to form cranked chains.

dopant ion.[1] As a result, the local electronic structure of the high-T_c superconductor is substantially different from the mean structure. This conclusion is central to our model [1].

In the scheme under consideration, a doped system comprising a strontium ion and a hole localized in the oxygen square is an electric dipole involved in long-range Coulomb interactions with the other similar dipoles. In such systems, the orientation interaction between dipoles appears; as a consequence, dipoles line up and their opposite poles are oriented toward each other. The crystal structure of the compound leads us to suggest that the substitution of strontium for lanthanum in $La_{2-x}Sr_xCuO_4$ occurs in such a way that the dipoles form chains (resembling crankshafts) along the c axis (Fig. 4b). This arrangement removes the question about strontium ions (from planes I, IV or II, III in Fig. 4a) that dope holes to the central plane.

We assume that the chains are planar and aligned with each other. Calculations of the electrostatic interaction energy between the chains of dipoles show that the nearest two chains (Fig. 4b) attract each other if the distance between doped holes is $l_{\text{com}} \geq \sqrt{2}$, while the interaction between the next-to-nearest chains is repulsive.[2] Such interactions provide ordering of dipole chains. As a result, doped holes in the CuO_2 plane (or projections of strontium ions) occur in square lattice sites with some parameter l_{com} commensurate with the lattice constant a in the CuO_2 plane. It follows from calculations that the energies of configurations with $l_{\text{com}} =$ 2, $\sqrt{5}$, $\sqrt{8}$, and 3 coincide to within $10^{-2}e^2/\varepsilon a$ per dipole, where ε is the permittivity. They can therefore coexist in the form of microdomains, in which doped holes occupy sites of lattices with various l_{com} values.

Domains with a given l_{com} distance can only exist over a certain concentration x range. This range is bounded from above by the $x_{\text{com}} = 1/l_{\text{com}}$ value; at higher concentrations, the existence of physically significant domains with given l_{com} violates the condition of a constant mean concentration. At $x < x_{\text{com}}$, dipole chains become broken, and vacancies appear in the square lattices of projections (Fig. 5). Microdomains with a given l_{com} distance remain intact up to some $x = x_l$ value, which, at a random distribution, corresponds to the two-dimensional percolation threshold on vacancies

---

[1] We can also consider another situation, when holes are doped into the central CuO_2 plane by strontium ions situated in planes II and III. It will, however, be shown that the reasoning presented below remains valid in this situation, and the final results remain unchanged.

[2] This corresponds to the experimental solubility limit of strontium impurity in La_2CuO_4 equal to $x_{\text{lim}} \approx 0.25$.

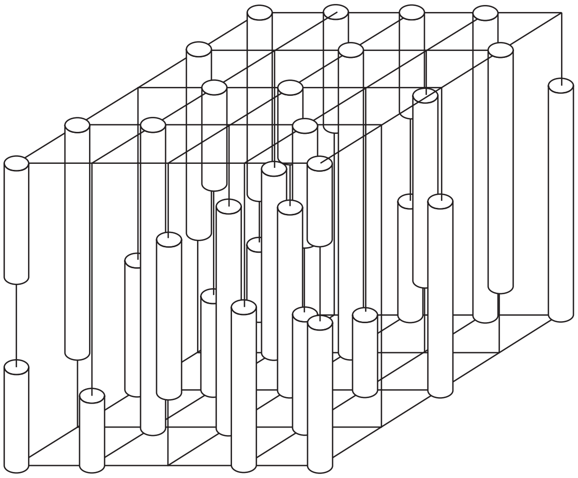


Fig. 5. Ordering of dipole chains at $x < x_{\text{com}}$. Columns are dipole chains aligned with the c axis. At $x < x_{\text{com}}$, dipole chains become broken, and vacancies appear in square lattices of the projections of strontium ions onto the CuO_2 plane.

with the concentration $c_v = 1 - x_l = 0.593$. Accordingly, the existence of domains with a given l_{com} value is possible at concentrations that satisfy the condition

$$\frac{0.407}{l_{\text{com}}^2} < x \le \frac{1}{l_{\text{com}}^2}. \qquad (1)$$

The concentration ranges in which domains with a given l_{com} value can exist are listed in the table.

At arbitrary concentrations x, domains of various types can coexist whose ordering corresponds to filling the sites of various projection lattices with l_{com} values that satisfy the condition

$$\frac{0.638}{\sqrt{x}} < l_{\text{com}} \le \frac{1}{\sqrt{x}}. \qquad (2)$$

For instance, at $x = 1/9$, domains with $l_{\text{com}} = 3$ (complete filling of the lattice of strontium ion projections) and $l_{\text{com}} = \sqrt{8}$, $\sqrt{5}$, and 2 can exist in volume ratios determined by the deviation of the given x value from x_{com} and the conditions of sample preparation. In what follows, the domain in which free square lattice sites with l_{com} are being filled will be characterized by the corresponding l_{com} value. The fraction of occupied sites changes from approximately 0.4 (at $x = 0.407/l_{\text{com}}^2$) to one (at $x = 1/l_{\text{com}}^2$) as x increases.

The ordered domain sizes depend on the concentration (more exactly, on the closeness of x to x_{com}) and reach 200–600 Å in the CuO_2 plane. The size of an ordered domain in the direction of the c axis is likely to be of several lattice constants because every type of ordering of doped charges is repeated in every second CuO_2 plane. Naturally, apart from domains with an ordered distribution of dopants, regions with disordered distribution must exist.

We assume that, at small x (at a mean distance between dopant projections of $l > 3$), dipole chains are grouped in planes parallel to the c axis and the orthorhombic a axis in such a way that the distance between doped holes (or strontium ion projections) along the a axis be $\sqrt{8}$; they that is, correspond to minimum interaction energy.

4. PERCOLATION AND THE T_c–x PHASE DIAGRAM OF $La_{2-x}Sr_xCuO_4$

As follows from our reasoning, $La_{2-x}Sr_xCuO_4$ must be treated as a set of mutually penetrating domains in which strontium ions are ordered in such a manner that doped holes fill (in part or completely) square lattice sites with various l_{com} values determined by the concentration.

The percolation regions on sites in lattices with various l_{com} values, that is, the concentration regions corresponding to the existence of continuous clusters of various phases, can be determined. The strontium concen-

Intervals of existence and site percolation thresholds for domains with various l_{com} values

l_{com}	x_0	x_p	x_m	Properties
>3				insulator
3	0.045	0.066	0.111	high-T_c superconductor (in the region of percolation)
$\sqrt{8}$	0.05	0.075	0.125	insulator
$\sqrt{5}$	0.08	0.12	0.20	high-T_c superconductor (in the region of percolation)
2	0.10	0.15	0.25	normal metal (in the region of percolation)

Note: x_0 and x_m are the lower and upper boundaries of the concentration range in which domains with the given l_{com} can exist, $x_0 = 0.407x_m$ and $x_m = 1/l_{\text{com}}^2$; $x_p = 0.593x_m$ is the percolation threshold for a lattice with period a, when the existence of percolation chains with $l = l_{\text{com}}$ becomes possible. The last column contains characteristics of microdomains with the given l_{com} value.

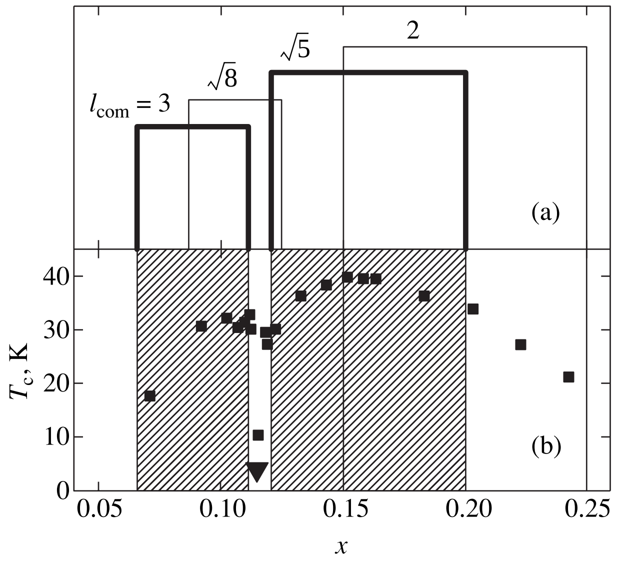


Fig. 6. (a) Concentration intervals corresponding to site percolation in domains with $l_{com} = 3, \sqrt{8}, \sqrt{5}$, and 2. Solid lines bound the regions of percolation broken lines with segments l_{com} long. Thick lines bound the regions of the existence of NUC percolation clusters. (b) Experimental T_c–x phase diagram for $La_{2-x}Sr_xCuO_4$ [18]. The composition ($x = 0.115$) for which superconductivity was not observed to 4.2 K is marked by a solid triangle.

tration x_p that corresponds to the percolation threshold is $x_p = 0.593/l_{com}^2$ for the given l_{com} at a random distribution over sites [14]. Figure 6a shows the intervals of concentrations corresponding to two-dimensional site percolation in domains with $l_{com} = 3, \sqrt{8}, \sqrt{5}, 2$, that is, the intervals where, according to our reasoning, two-dimensional percolation clusters with $l_{com} = 3$ and $\sqrt{5}$ (NUC chains), a cluster with $l_{com} = 2$ corresponding to the usual metal phase, and a cluster with $l_{com} = \sqrt{8}$ corresponding to the insulator phase can exist. The boundaries of the regions of the existence of percolation broken lines with segment lengths l_{com} are shown by solid lines in Fig. 6a. The percolation regions for domains with $l_{com} = 3$ and $\sqrt{5}$, that is, for NUC chains, are indicated by thick lines. Figure 7 shows how the regions of the localization of singlet hole pairs overlap along percolation clusters with (a) $l_{com} = 3$ and (b) $l_{com} = \sqrt{5}$. Note that, as is seen from Fig. 7a, current transfer largely occurs along Cu–O bonds if $l_{com} = 3$. This is in agreement with the results reported in [16, 17], where the conclusion of such a character of the movement of carriers was drawn from the ARPES, IR, and Raman data on $La_{1.9}Sr_{0.1}CuO_4$ crystals.

Figure 6a shows that volume superconductivity (domains with percolation NUC chains) exists in the regions $0.066 < x < 0.11$ and $0.12 < x < 0.20$. Superconducting domains (containing percolation clusters with $l_{com} = \sqrt{5}$) and normal metal domains coexist at $0.15 < x < 0.20$; the fraction of normal domains increases as x grows. This corresponds to the transition to the state in which superconductivity is fully determined by the neighborhood effect and T_c monotonically decreases as x increases.

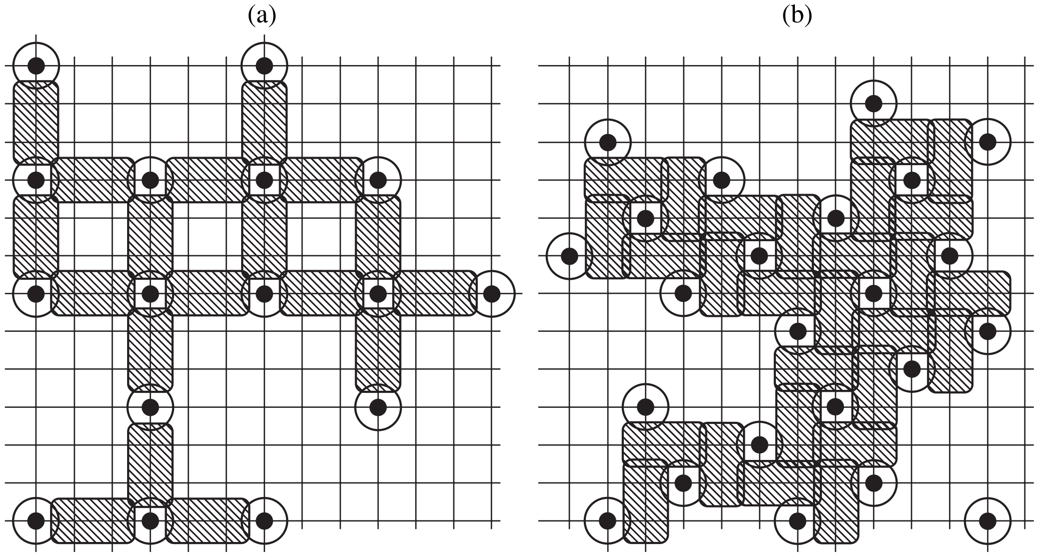


Fig. 7. Percolation clusters in the regions of the localization of singlet hole pairs: (a) $l_{com} = 3$ and (b) $l_{com} = \sqrt{5}$. Solid circles are the projections of dopant ions onto the CuO_2 plane, open circles are doped hole localization regions, and rectangles are the localization regions of hole pairs around NUCs.

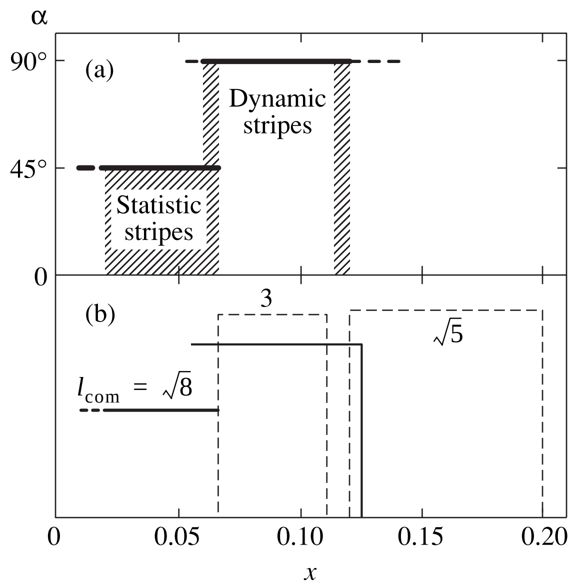


Fig. 8. (a) Experimental magnetic phase diagram of $La_{2-x}Sr_xCuO_4$ [11, 32–36]; $\alpha = 45°$ and 90° correspond to diagonal and vertical stripes, respectively. Hatched regions are the intervals where static stripes are observed. (b) Calculated stripe phase diagram of $La_{2-x}Sr_xCuO_4$. The dashed lines bound the regions of percolation along NUC chains with $l_{com} = 3$ and $l_{com} = \sqrt{5}$ (dynamic stripes); the thick lines correspond to the regions of the existence of microdomains with doped holes ordered into a $\sqrt{8} \times \sqrt{8}$ lattice ($0.05 < x < 0.12$) and diagonal lines of doped holes ($x < 0.066$).

For comparison, the experimental T_c–x phase diagram of $La_{2-x}Sr_xCuO_4$ [18] is shown in Fig. 6b. The coincidence of the superconductivity intervals in the experimental phase diagrams and the percolation intervals for $l_{com} = \sqrt{5}$ and $l_{com} = 3$ substantiates the conclusion that it is the fragments under consideration including pairs of neighboring copper ions that are responsible for the superconductivity of $La_{2-x}Sr_xCuO_4$. This is also evidence in favor of the suggested model of high-temperature superconductivity. Note that the "dip" in the T_c–x diagram at $0.11 < x < 0.12$, which is related to the absence of percolation along NUC chains, is superimposed on the narrow region of the existence (for $x \longrightarrow 1/8$) of the fully ordered $\sqrt{8} \times \sqrt{8}$ lattice of doped charges, which corresponds to the insulator phase. We show in the next section that precisely this allows us to observe a statistical incommensurate magnetic texture in this region [11].

As follows from the above analysis, the microstructure of $La_{2-x}Sr_xCuO_4$ must be treated as a set of NUC clusters of various sizes immersed into either a insulator (at $x < 0.125$) or a metal-insulator (at $0.125 < x < 0.20$) matrix. A conducting cluster at $x < 0.2$ can therefore include insulator regions whose conductivity is determined exclusively by quantum tunneling processes in combination with transfer along the c axis. The true metallic conductivity is therefore only possible when $x \longrightarrow 0.2$. This corresponds with the results obtained in [19], where a logarithmic divergence of resistance was observed at $x < 0.17$ as T tended to zero when superconductivity was suppressed by a magnetic field.

To summarize, the conclusion can be drawn that all the special features of the T_c–x phase diagrams of $La_{2-x}Sr_xCuO_4$ only reflect the geometric relations that exist in a square lattice and competition between various strontium ordering types.

5. INCOMMENSURATE CHARGE AND SPIN MODULATION IN $La_{2-x}Sr_xCuO_4$

In recent years, the concept of stripes has been used in one or another form to analyze the results obtained in many works concerned with hole-doped cuprate high-T_c superconductors [20–31]. This concept presupposes the existence of incommensurate modulation of the spin antiferromagnetic structure in the form of antiphase domains of antiferromagnetically ordered spins separated by narrow extensive stripes of doped holes.

In experiments on the magnetic scattering of neutrons, such a modulation characterized by the wavevector **Q** should be observed in the form of two incommensurate peaks shifted with respect to the antiferromagnetic wavevector $\mathbf{Q}_{AF}(1/2, 1/2, 0)$ by $\varepsilon = 1/T$ in the direction of the modulation vector. Here, T is the magnetic structure period in lattice constant units. Accordingly, the charge modulation period should be $T/2$, and the related incommensurate charge density modulation should manifest itself in the appearance of additional peaks shifted by 2ε.

The neutron diffraction data on the magnetic texture of $La_{2-x}Sr_xCuO_4$ and $La_{1.6-x-y}Nd_{0.4}Sr_xCuO_4$ [11, 32–36] can be summarized in the form of the phase diagram shown in Fig. 8a. According to this figure, the incommensurate elastic scattering peaks related to statistical modulation (hatched in the figure) are observed at strontium concentrations $x < 0.07$ and $0.11 < x < 0.12$. In the interval $0.07 < x < 0.11$ and at $x > 0.13$, incommensurate inelastic neutron scattering peaks appear, which is evidence of dynamic modulation of the spin texture. At $x < 0.07$, there are diagonal stripes with a single modulation vector directed along the orthorhombic b axis, whereas for parallel stripes ($x > 0.055$), there is always modulation in two directions parallel to the tetragonal axes. In the intermediate region $0.055 < x < 0.07$, both modulation types are observed. To compare the spin structures that appear when there are diagonal and parallel stripes, both are considered in the tetragonal coordinates. The incommensurability parameter of spin modulation is then $\delta = \varepsilon$ for parallel and $\delta = \varepsilon/\sqrt{2}$

for diagonal stripes. For $x < 0.12$, we experimentally observe a remarkably simple relation between the incommensurability parameter and concentration, namely, $\delta \approx x$.

The appearance of a stripe structure caused by the competition between electronic phase separation and long-range Coulomb repulsion was studied theoretically in [20–27]. An alternative mechanism of the formation of the insulator “stripe” phase in an underdoped high-T_c superconductor caused by Fermi surface nesting was suggested in [28–31]. Both theories, however, face difficulties in describing the whole complex of experimental data. The most important of these are:

(1) the transition from diagonal to parallel stripes in $La_{2-x}Sr_xCuO_4$ at $x \approx 0.06$;

(2) the appearance of static correlations in $La_{2-x}Sr_xCuO_4$ in a narrow concentration range at $x \approx 0.12$;

(3) the one-dimensional character of diagonal stripes and the two-dimensional character of parallel stripes;

(4) the $\delta \approx x$ relation for $x \leq 0.12$ and $\delta = \text{const}$ at $x \geq 0.12$. As follows from Hall effect measurements, the x value in $La_{2-x}Sr_xCuO_4$ is not equal to the concentration of hole carriers (which, in addition, strongly depends on temperature). The equality $\delta \approx x$ is therefore indicative of a relation between spin texture parameters and the concentration of strontium ions rather than the density of carriers.

In attempts to overcome the difficulties that arise when the neutron scattering data are analyzed, Gooding *et al.* [37, 38] suggested a physical spin glass model based on the assumption of a chaotic distribution of localized doped holes. According to [38], doped holes in the spin glass phase are localized in the CuO_2 plane. Such a hole localized in a certain region creates a long-range field of spin distortions around it. The appearing distortions of the antiferromagnetic background can be described as the creation of a topological excitation similar to skyrmion [39, 40] with topological charges $Q = \pm1$. This excitation corresponds to rotation (twist) of the antiferromagnetic order parameter (Fig. 9) in the vicinity of a localized hole.

Doping therefore destroys long-range antiferromagnetic order and results in the formation of disoriented antiferromagnetically ordered microdomains. The boundaries between them (domain walls) are determined by localized doped holes, and the directions of antiferromagnetic ordering in the neighboring domains are rotated through some angle with respect to each other (so-called spin twisting). This model allows various special features of the spin texture of $La_{2-x}Sr_xCuO_4$ observed in the spin glass state to be explained [41].

In what follows, we give an alternative explanation of the observed spin and charge modulation. Certain presumptions of the model suggested in [37, 38] will be combined with the concepts of the mechanism of formation of NUCs, dopant ordering, and percolation developed in this work. For this purpose, let us first try to guess the form of the spin texture (different from the classic stripe picture) for some strictly ordered distributions of doped holes with the concentration $x = x_{com}$, which is responsible for the experimentally observed spin modulation picture. This approach is justified, because there are no grounds to expect that the stripe model is inapplicable when the distribution is ordered. Moreover, it follows from the experiment reported in [36] that, in the region $0.06 < x < 0.12$ (the region of parallel stripes), the correlation length increases from 25 Å at $x \approx 0.06$ to 200 Å at $x \approx 0.12$. This increase in the correlation region size is natural to relate to ordering of separate antiferromagnetic domains. It can be suggested that this ordering is in turn related to the ordering of doped holes and, accordingly, dopant ions.

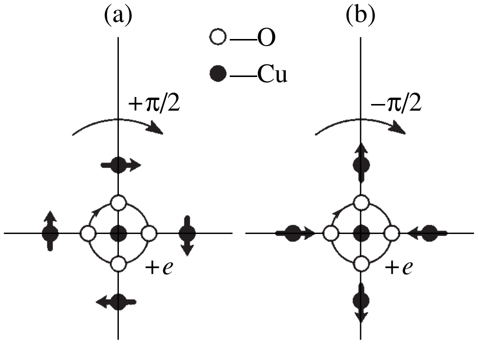


Fig. 9. Rotations of copper ion spin directions (thick arrows) in the vicinity of a doped hole localized in the CuO_2 plane on four oxygen atoms around the copper ion; rotations correspond to different skyrmion topological signs (a) $Q = 1$ and (b) $Q = -1$.

Next, we will consider how the texture guessed by us transforms as x decreases below x_{com} and as the system deviates from the strictly ordered distribution of holes. We will show that, in a certain concentration region $x_p < x < x_{com}$, the principal experimentally observed relations that are characteristic of an ideal lattice of holes ($x = x_{com}$) remain unchanged.

5.1. Parallel Stripes

Let us consider complete ordering at $x_{com} = 1/8$ assuming that (1) each hole circulates over the oxygen square that surrounds a copper ion, (2) because of the interaction between the hole current and the spins of the nearest four copper ions, these ions are polarized, and the resulting distortions of the antiferromagnetic background can be described as the creation of a skyrmion with topological charges $Q = \pm1$ (Fig. 9).

Possible ordering of the projections of copper spin directions onto the CuO_2 plane is shown in Fig. 10a for a completely ordered arrangement of localized holes at $x = x_{com} = 1/8$, when the holes form a $\sqrt{8} \times \sqrt{8}$ square

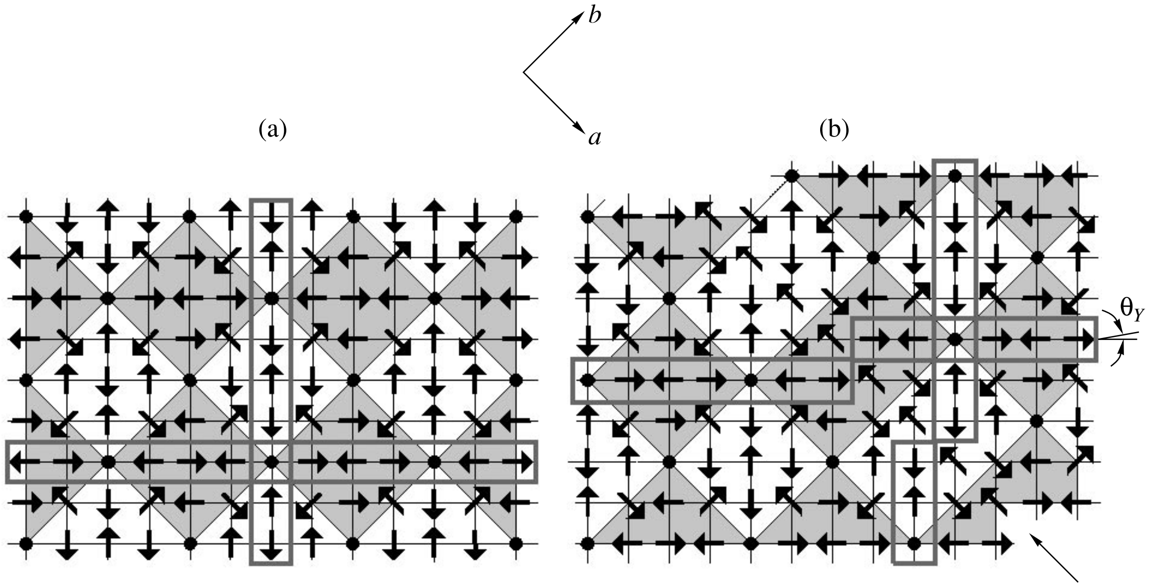


Fig. 10. (a) Projections of spin directions at $x = 1/8$ when doped holes are ordered into a $\sqrt{8} \times \sqrt{8}$ lattice. Microdomains that form horizontal stripes are hatched, thick lines denote stripe directions. (b) The same at $x < 1/8$. The plane is divided into domains separated by diagonal dislocations, which are nuclei of diagonal stripes. The shift of vertical stripes by one cell at each dislocation results in an effective tilt of vertical stripes by the θ_Y angle.

lattice. The CuO_2 plane is then divided into separate quadrangular antiferromagnetically ordered microdomains; the quadrangle angles are determined by the localized doped holes. The projections onto the CuO_2 plane of the spin directions of the copper ions in lattice sites are shown by arrows. The correlated ordering that then appears is characterized by the antiferromagnetic ordering of microdomains themselves and the ordered alternation of skyrmions with charges $Q = \pm1$. As is seen from Fig. 10a, this ordering simulates a magnetic stripe texture. The magnetic modulation period is then equal to the sum of the sizes of two antiphase domains in the direction of the modulation vector. A stripe of width $l_{\text{com}}/\sqrt{2}$ and length equal to one period then contains two sites,

$$T = 2\sqrt{2}l_{\text{com}} = 8, \tag{3}$$

and $\delta = 1/8 = x$, in agreement with experiment. This picture is in conformity with what was observed in [42], where μsR relaxation measurements revealed the existence of antiferromagnetically ordered microdomains in $La_{1.88}Sr_{0.12}CuO_4$. These microdomains were 15–30 Å in size, and magnetization directions in them were correlated on scales up to 600 Å. Note that, as follows from Fig. 8a, there are no charge stripes in this system in the form predicted theoretically. However, we here deal with commensurate modulation, which cannot give satellite diffractogram peaks. Satellite reflections can only appear when modulation is incommensurate. In addition, charge modulation was also observed in [11] in the form of the incommensurate lattice peak splitting $(2 \pm 2\varepsilon, 0, 0)$ and $(0, 2 \pm 2\varepsilon, 0)$.

To find the reason for this, let us pass from the fully ordered lattice of doped holes at $x = 1/8$ to the distribution of holes at $x < 1/8$. Let us first consider the experimental data obtained in the region of the existence of static parallel stripes at $x \approx 0.12$ [36, 43]. This will allow us to compare the experimental results with our model in detail.

Spin texture modulation with the incommensurability parameter $\delta = 0.118$ was observed in [43] for $La_{1.88}Sr_{0.12}CuO_4$. This corresponds to the mean texture period $T \approx 8.5$ (in a units), that is, to the alternation of two periods $T_1 = 8$ and $T_2 = 9$. The picture that we suggest to describe an ordered distribution of doped holes (therefore, dopant projections) for the mean concentration $x = 0.118$ is shown in Fig. 10b. This picture was obtained by cutting the fully ordered distribution ($x = 0.125$) along the orthorhombic axis a and shifting one part with respect to the other by the $\mathbf{q} = (1, 1)$ vector. This shift does not disturb the coherence of ordering in domains on both sides of the dislocation and shifts the system of parallel stripes by one cell in traversing the cut.[3] Such a structure (Fig. 10b) gives characteristic diffractogram reflections corresponding to incommensurate modulation of both spin (with the incommensurability parameter δ) and charge (incommensurability

---

[3] We in fact suppose that dipole chains (therefore, vacancy chains) tend to line up along the orthorhombic axis a. This is, in our view, caused by the minimum strain requirement.

parameter 2δ). The condition of a constant mean concentration yields

$$T_d x_l = (T_d + 1)\bar{x}, \qquad (4)$$

where T_d is the mean period of dislocations in *a* units and x_l is the local concentration of holes inside a domain. To satisfy the condition of the constant mean concentration $\bar{x} = 0.118$, the local concentration inside domains being $x_l = 0.125$, the diagonal dislocations that we introduce should have the mean period $T_d = T_1 + T_2 = 17$. Such quasi-periodic dislocations resulting in incommensurate modulation of both the crystalline and spin textures are responsible for the possibility of observing incommensurate reflections in diffraction experiments.

The special feature of the appearing picture of ordering is the shift of parallel stripes by one lattice constant (see Fig. 10b); that is, it is as though these stripes are tilted by the $\theta_Y = 1/17 \approx 3.3°$ angle with respect to the tetragonal *b* axis. It is these tilted parallel stripes with a 3° tilt angle that were observed in $La_{1.88}Sr_{0.12}CuO_4$ [43].

Next, let us consider arbitrary $x < 1/8$ values. The distribution picture can then be obtained from the fully ordered distribution at $x = 1/8$ by successively removing a certain number of sites. A texture simulating parallel stripes can appear if there exists a percolation cluster that binds $\sqrt{8} \times \sqrt{8}$ microdomains into one antiferromagnetically ordered cluster. The table shows that such microdomains can exist up to $x \approx 0.05$.

Let the lattice contain correlated residual fragments of a parallel stripe texture genetically related to $\sqrt{8} \times \sqrt{8}$ microdomains (Fig. 11). The neutron diffraction pattern then contains characteristic reflections determined by the mean period of the residual texture. The mean period T of this texture, which is the distance between the middle points of single-phase magnetic microdomains, includes two occupied sites, as when there is complete ordering. It follows that the rectangle of area $Tl_{\text{com}}/\sqrt{2} = 2T$ (this rectangle is hatched in Fig. 11) should contain two sites. Therefore,

$$2Tx = 2, \quad \delta = 1/T = x. \qquad (5)$$

It follows that the $\delta = x$ equality is in some sense fortuitous. It exists because, in the case of parallel stripes, holes lie on straight lines situated at distances $2a$ from each other.

5.2. Diagonal Stripes

Figure 10a shows that the dislocations introduced are in fact nuclei of diagonal stripes extended along the orthorhombic axis *a*. They appear as a quasi-periodic structure at $x < 0.05$, when $\sqrt{8} \times \sqrt{8}$ texture remainders disappear and there only remain diagonal lines of impurity dipoles with a distance of $l_{\text{com}} \geq 2\sqrt{8}$ between the lines and a distance of $l_{\text{com}} = \sqrt{8}$ between the dipoles. Diagonal stripes are therefore always directed along the orthorhombic axis *a*, and, accordingly, the modulation vector, along the other orthorhombic axis *b*.

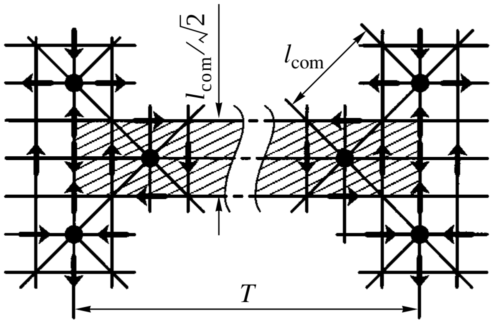


Fig. 11. Fragment of magnetic stripe texture. Solid circles are the projections of strontium ions onto the CuO_2 plane. Arrows show the directions of spin projections at Cu sites. Two filled sites occur within a band of width $l_{\text{com}}/\sqrt{2}$ over the length T of the magnetic structure period (the rectangle of area $Tl_{\text{com}}/\sqrt{2}$ is hatched).

The period T of diagonal spin modulation (in tetragonal axes) should be $T = 1/\sqrt{2}x$ (or $\delta = \sqrt{2}x$) when all doped holes are ordered into diagonal charge stripes. Since part of the doped holes can remain in the space between charge stripes, the period of the observed spin texture is larger than $1/\sqrt{2}x$ and, accordingly, δ is smaller than $\sqrt{2}x$. The experimental δ values vary from $\delta \approx 0.7x$ to $\delta \approx 1.4x$ at $0.01 < x < 0.05$ [34].

5.3. Dynamic Stripes

The last question discussed in this Section concerns static and dynamic stripes. Figure 8b shows concentration regions where antiferromagnetically correlated clusters of $\sqrt{8} \times \sqrt{8}$ microdomains and diagonal lines of dipoles can exist. Primes denote regions of the existence of percolation chains with $l_{\text{com}} = \sqrt{5}$ and $l_{\text{com}} = 3$. Such chains of doped holes in the CuO_2 plane can border on a cluster of antiferromagnetic microdomains. According to our model, NUCs are formed in regions of the existence of percolation clusters with $l_{\text{com}} = 3$ and $l_{\text{com}} = \sqrt{5}$ on pairs of neighboring copper ions; these NUCs play the role of pair acceptors. In these regions, we observe conductivity along the corresponding NUC chains situated on percolation clusters. The appearance of conductivity disturbs static spin correlations in the surrounding regions because of the movement of charges that destroy magnetic order along their trajectories. Spin correlations can then be observed only in inelastic neutron scattering as dynamic incommensurate magnetic fluctuations. Remarkably (see Fig. 8b), apart from the region $x < 0.07$, there is a narrow concentration range $0.11 < x < 0.125$ where NUC percolation is absent. This is where static incommensurate correlations are again observed.

6. CONCLUSIONS

In this work, we suggested a microstructural model of the $La_{2-x}Sr_xCuO_4$ compound, which, in combination with the results reported in our work [1], allowed us to qualitatively and quantitatively explain all the details of the T_c–x phase diagram of $La_{2-x}Sr_xCuO_4$, in particular:

(1) spatial inhomogeneity on the nanoscopic scale observed experimentally;

(2) the existence of two concentration ranges of volume high-temperature superconductivity ($0.066 < x < 0.11$ and $0.12 < x < 0.2$);

(3) the $T_c(x)$ curve dip at $x \approx 1/8$;

(4) successive transition from underdoping conditions (at $x < 0.12$, where percolation clusters of negative U-centers are “immersed” into a insulator matrix) to optimum doping ($x \approx 0.15$) and further to overdoping conditions (at $x > 0.15$, where percolation clusters of negative U-centers coexist with usual metal clusters);

(5) the superconductor–insulator transition in a magnetic field in underdoped samples;

(6) a gradual decrease in T_c to zero as x increases in the overdoping region (at $x > 0.15$).

To check the validity of the suggested model of ordering, we considered magnetic spin textures, which can be observed if the character of ordering of doping ions corresponds to that suggested in this work. We found that the spin textures that are formed completely simulate so-called stripe modulation (electronic phase separation into charge and antiferromagnetic insulator stripes) earlier believed to be responsible for the appearance of incommensurate reflections in magnetic neutron diffractograms.

On the other hand, we were able to explain the transition from diagonal to vertical and from static to dynamic stripes and back. The predicted concentration regions, where diagonal ($x < 0.05$), vertical ($0.05 < x < 0.12$), and dynamic ($0.066 < x < 0.11$) stripes should exist according to our model, coincide exactly with the regions determined experimentally.

It follows from our model, in qualitative and quantitative agreement with experiment, that tilted stripes with tilt angles depending on the incommensurability parameter δ can be observed. It also naturally follows from this model that the incommensurability parameter δ is related to the concentration as $\delta = x$. This equality is only valid in the region $0.05 < x < 0.12$, in agreement with the experimental data.

The suggested model describes the experimental spin textures in detail. This is evidence that the model is valid for La_2CuO_4 and therefore applicable to analyzing the T_c–x phase diagram. On the other hand, both (“superconducting” and “magnetic”) phase diagrams were constructed using not only the ordering model but also the concept of negative U-centers developed by us earlier. The coincidence of the experimental and calculated phase diagrams can therefore be considered evidence of the validity of the assumptions made and of a key role played by negative U-centers of the suggested type in the mechanism of high-temperature superconductivity.

ACKNOWLEDGMENTS

The authors thank E.G. Maksimov, P.I. Arseev, and the participants of the Workshop on the Theory of Solids (Division of Theoretical Physics, Lebedev Physical Institute, Russian Academy of Sciences) for useful discussions. This work was financially supported by the Federal Agency of Science of the Russian Federation (Scientific-Technical Program “Theoretical and Experimental Studies of the Mechanisms of High-Temperature Superconductivity”).

REFERENCES

1. K. V. Mitsen and O. M. Ivanenko, Zh. Éksp. Teor. Fiz. **118**, 666 (2000) [JETP **91**, 579 (2000)].
2. E. Simanek, Solid State Commun. **32**, 731 (1979).
3. C. S. Ting, D. N. Talwar, and K. L. Ngai, Phys. Rev. Lett. **45**, 1213 (1980).
4. H.-B. Schuttler, M. Jarrell, and D. J. Scalapino, Phys. Rev. Lett. **58**, 1147 (1987).
5. J. Yu, S. Massida, A. J. Freeman, *et al.*, Phys. Lett. A **122**, 203 (1987).
6. B. A. Volkov and V. V. Tugushev, Pis’ma Zh. Éksp. Teor. Fiz. **46**, 193 (1987) [JETP Lett. **46**, 245 (1987)].
7. G. M. Éliashberg, Pis’ma Zh. Éksp. Teor. Fiz. **46** (Suppl.), 94 (1987) [JETP Lett. **46**, S81 (1987)].
8. I. O. Kulik, Fiz. Nizk. Temp. **13**, 879 (1987) [Sov. J. Low Temp. Phys. **13**, 505 (1987)].
9. P. I. Arseev, Zh. Éksp. Teor. Fiz. **101**, 1246 (1992) [Sov. Phys. JETP **74**, 667 (1992)].
10. J. Ranninger and A. Romano, Phys. Rev. B **66**, 94508 (2002).
11. J. M. Tranquada, J. D. Axe, N. Ichikawa, *et al.*, Phys. Rev. B **54**, 7489 (1996).
12. V. Polinger, D. Haskel, and E. A. Stern, AIP Conf. Proc. **483**, 241 (1999).
13. P. C. Hammel, B. W. Statt, R. L. Martin, *et al.*, Phys. Rev. B **57**, R712 (1998).
14. R. M. Ziff, Phys. Rev. Lett. **69**, 2670 (1992).
15. R. Beyers, B. T. Ahn, G. Gorman, *et al.*, Nature **340**, 619 (1989).
16. A. Ino, C. Kim, M. Nakamura, *et al.*, Phys. Rev. B **65**, 094504 (2002).
17. F. Venturini, Q.-M. Zhang, R. Hackl, *et al.*, Phys. Rev. B **66**, 060502 (2002).

18. K. Kumagai, K. Kawano, I. Watanabe, *et al.*, J. Supercond. **7**, 63 (1994).
19. Y. Ando, G. S. Boebinger, A. Passner, *et al.*, Phys. Rev. B **56**, R8530 (1997).
20. J. Zaanen and O. Gunnarson, Phys. Rev. B **40**, 7391 (1989).
21. D. Poilblanc and T. M. Rice, Phys. Rev. B **39**, 9749 (1989).
22. S. A. Kivelson, V. J. Emery, and H.-Q. Lin, Phys. Rev. B **42**, 6523 (1990).
23. H. Schulz, Phys. Rev. Lett. **64**, 1445 (1990).
24. M. Inui and P. Littlewood, Phys. Rev. B **44**, 4415 (1991).
25. G. An and J. M. van Leeuwen, Phys. Rev. B **44**, 9410 (1991).
26. V. J. Emery and S. A. Kivelson, Physica C (Amsterdam) **209**, 597 (1993).
27. C. C. Castellani, C. Di Castro, and M. Grilli, Phys. Rev. Lett. **75**, 4650 (1995).
28. Q. Si, Y. Zha, K. Levin, and J. P. Lu, Phys. Rev. B **47**, 9055 (1993).
29. T. Tanamoto, H. Kohno, and H. Fukuyama, J. Phys. Soc. Jpn. **63**, 3739 (1994).
30. R. S. Markiewicz, Phys. Rev. B **56**, 9091 (1997).
31. V. I. Belyavsky and Y. V. Kopaev, Phys. Lett. A **287**, 152 (2001).
32. K. Yamada, C. H. Lee, K. Kurahashi, *et al.*, Phys. Rev. B **57**, 6165 (1998).
33. K. Yamada, C. H. Lee, Y. Endoh, *et al.*, Physica C (Amsterdam) **282–287**, 85 (1997).
34. M. Matsuda, M. Fujita, K. Yamada, *et al.*, Phys. Rev. B **65**, 134515 (2002).
35. M. Fujita, K. Yamada, H. Hiraka, *et al.*, Phys. Rev. B **65**, 64505 (2002).
36. M. Fujita, H. Goka, K. Yamada, *et al.*, Phys. Rev. B **66**, 184503 (2002).
37. R. J. Gooding, N. M. Salem, and A. Mailhot, Phys. Rev. B **49**, 6067 (1994).
38. R. J. Gooding, N. M. Salem, R. J. Birgeneau, *et al.*, Phys. Rev. B **55**, 6360 (1997).
39. A. A. Belavin and A. M. Polyakov, Pis'ma Zh. Éksp. Teor. Fiz. **22**, 503 (1975) [JETP Lett. **22**, 245 (1975)].
40. R. J. Gooding, Phys. Rev. Lett. **66**, 2266 (1991).
41. R. J. Birgeneau, A. Cassanho, C. Y. Chen, *et al.*, Phys. Rev. B **46**, 14034 (1992).
42. A. T. Savici, Y. Fudamoto, I. M. Gat, *et al.*, Phys. Rev. B **66**, 14524 (2002).
43. H. Kimura, H. Matsushita, K. Hirota, *et al.*, Phys. Rev. B **61**, 14366 (2000).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 100, No. 6, 2005, pp. 1094–1111.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 6, 2005, pp. 1244–1261.
Original Russian Text Copyright © 2005 by Mirsaev, Turov.

ORDER, DISORDER, AND PHASE TRANSITIONS IN CONDENSED SYSTEMS

Electron and Nuclear Magnetic Resonances Associated with Magnetoelectric and Antiferroelectric Interactions in the Exchange-Noncollinear (“Square”) Chiral Antiferromagnet Nd_2CuO_4 and Their Behavior in Chiral Phase Transitions

I. F. Mirsaev and E. A. Turov

Institute of Metal Physics, Ural Division, Russian Academy of Sciences, Yekaterinburg, 620219 Russia

e-mail: mirsaev@imp.uran.ru; turov@imp.uran.ru

Received July 12, 2004

Abstract—Experiments with the tetragonal antiferromagnet Nd_2CuO_4 in the temperature range 1.5 K < T < T_N = 245 K show that the magnetic moments of Cu^{2+} possess an exchange-noncollinear magnetic structure of the “square” type, which has the form of an exchange doublet whose components exhibit different chiralities (Γ_4 and Γ_5 phases). Between these phases, consecutive phase transitions $\Gamma_4 \longleftrightarrow \Gamma_5 \longleftrightarrow \Gamma_4$ with a change in chirality take place at temperatures T_1 = 30 K and T_2 = 70 K. The electron and nuclear magnetic resonances (natural frequencies and susceptibilities) associated with excitation of magnons (due to the magnetoelectric and antiferroelectric interactions) by an ac electric field $\mathbf{E}(t)$, as well as a variable magnetic field $\mathbf{H}(t)$ applied in the case of a constant electric field $\mathbf{E}^0$, are calculated. It is predicted that nuclear magnetic resonance is excited by an ac electric field at frequencies determined by hyperfine fields of the sublattices. The change in the resonance frequencies upon the above chiral phase transitions are analyzed (being first-order phase transitions, these transitions possess a number of features associated with the chirality of the magnetic structures). *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Rare-earth cuprates R_2CuO_4 (R = Nd, Pr, Sm, etc.) have attracted the attention of researchers owing to their possible use for preparing high-temperature superconductors (by adding certain impurities). However, it turned out that these compounds themselves are of considerable interest as regards their magnetic (usually antiferromagnetic) properties. This can be demonstrated using neodymium cuprate (Nd_2CuO_4) as an example, for which numerous theoretical and experimental studies were carried out and clearly formulated results were obtained. From the very outset, considerable difficulties and inconsistencies in the crystalline structure of Nd_2CuO_4 determined by different authors have been observed. At $T > T_c$ = 300 K, in the paramagnetic region, a consistent result was obtained, according to which the structure is determined by the body-centered tetragonal group

$$I4/mmm \quad (D_{4h}^{17}); \tag{1}$$

however, in the antiferromagnetic ordered region $T < T_N$ = 245 K, a substantial disagreement between two groups of publications is observed. According to the results obtained in the first group [1–4], the same crystal-chemistry symmetry group (1) is preserved in the antiferromagnetic region, in which a primitive cell contains two Cu^{2+} ions in view of body centering, so that antitranslation $\tau' = \tau(1/2, 1/2, 1/2) \times 1'$ appears in the magnetic groups as a result of antiferromagnetic ordering [5] (1' denotes the time inversion $t \longrightarrow -t$). According to the second group of authors [6–9], a structural phase transition occurs at T_c = 300 K to the group

$$P4_2/mnm \quad (D_{4h}^{14}). \tag{2}$$

Although this group is also tetragonal, it is simple and noncentered; the unit cell volume of this group increases by a factor of 4 due to doubling of each edge of the unit cell in the basal plane. In this case, the antitranslations typical of the preceding group (1) do not appear in the antiferromagnetic structure.

Our analysis is based on the assumption that the phase transition from group (1) to group (2) indeed takes place, although, as mentioned above, the existence of this transition is not recognized by some authors. We present some arguments in favor of this transition and develop the spin-wave dynamics for

Nd_2CuO_4 taking into account the magnetoelectric and antiferroelectric interactions taking place in the region of antiferromagnetic ordering of magnetic moments of copper ions (i.e., at quite high temperatures of $T \geq 1.5$ K, for which the ordering of the magnetic moments of neodymium ions does not take place). We will consider antiferromagnetic structures and specific (chiral) phase transitions between them, which take place in the above-mentioned temperature range.

2. ON MAGNETIC STRUCTURE OF Nd_2CuO_4 ACCORDING TO EXPERIMENTAL DATA: CHIRALITY

Thus, following the authors of [6–9], we assume that the antiferromagnetic compound Nd_2CuO_4 (T_N = 245 K) has a tetragonal crystal lattice with space symmetry $P4_2/mnm$ (D_{4h}^{14}) [7]. In a unit cell (Fig. 1), the Cu^{2+} magnetic ions occupy the fourfold position $4f\{mm\}$:

$$1(x, x, 0), \quad 2(1/2 + x, 1/2 - x, 1/2),$$
$$3(1/2 - x, 1/2 + x, 1/2), \quad 4(1 - x, 1 - x, 0).$$

The local symmetry of the Cu^{2+} ions is indicated in the braces. Four magnetic sublattices with magnetizations $\mathbf{M}_\nu$ (ν = 1, 2, 3, 4) correspond to four (one ferromagnetic and three antiferromagnetic) vectors:

$$\begin{aligned} \mathbf{M} &= \mathbf{M}_1 + \mathbf{M}_2 + \mathbf{M}_3 + \mathbf{M}_4, \\ \mathbf{L}_a &= \mathbf{M}_1 + \mathbf{M}_2 - \mathbf{M}_3 - \mathbf{M}_4, \\ \mathbf{L}_b &= \mathbf{M}_1 - \mathbf{M}_2 + \mathbf{M}_3 - \mathbf{M}_4, \\ \mathbf{L}_c &= \mathbf{M}_1 - \mathbf{M}_2 - \mathbf{M}_3 + \mathbf{M}_4. \end{aligned} \tag{3}$$

We can compose a table (Table 1) of transformations of these vectors under the action of generators of group $P4_2/mnm$, assuming that elements $\bar{1}$, 2_{1x}, and 4_{2z} are the end elements (see Fig. 1). It should be borne in mind that these elements not only generate conventional rotations and reflections, but also carry out the transposition of copper ions (space symmetry!). It can be seen from Table 1 that two of four vectors (3) ($\mathbf{M}$ and $\mathbf{L}_c$) are centrosymmetric, so that

$$\bar{1}\mathbf{M} = \mathbf{M} \quad \text{and} \quad \bar{1}\mathbf{L}_c = \mathbf{L}_c \tag{4}$$

(where $\bar{1}$ is the space inversion operator), while the remaining two operators are centroantisymmetric:

$$\bar{1}\mathbf{L}_{a,b} = -\mathbf{L}_{a,b}. \tag{5}$$

Pay attention to the fact that Table 1 is compiled so that it first represents the transformations for the rhombic group $Pnmm$ (column 3), which is a subgroup of the $P4_2/mnm$ group we are interested in since $4_z^2 = 2_z$. The first column labels phases Γ_n of the $Pnmm$ group in terms of the components of vectors $\mathbf{M}$ and $\mathbf{L}_\xi$ ($\xi = a, b, c$). Numbers +1 and –1 in column 3 indicate whether the corresponding functions from the second column reverse their sign under the action of elements $\bar{1}$, 2_{1x}, and 2_z. The last column contains the results of transformation of these functions under the action of element 4_{2z} supplementing the $Pnmm$ group to $P4_2/mnm$.

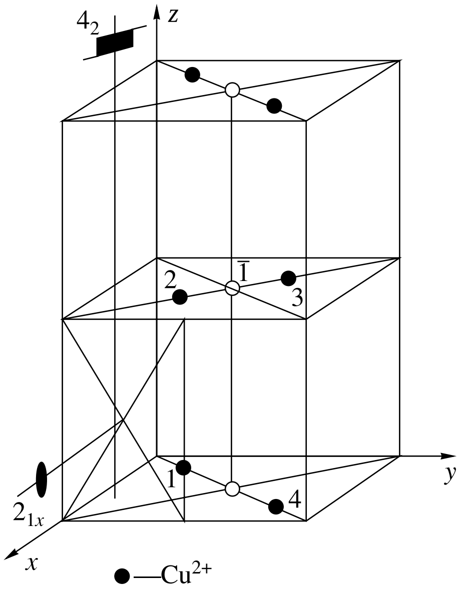


Fig. 1. Unit cell of the crystal lattice of Nd_2CuO_4 for the $4f$ position of the $P4_2/mnm$ (D_{4h}^{14}) group. Positions of generators $\bar{1}$, 2_{1x}, and 4_{2z} of the group are indicated.

It should be noted from the very outset that the magnetic ions Cu^{2+} occupy noncentrosymmetric positions so that spatial inversion $\bar{1}$ transposes atoms with one another and not each atom into itself; this ensures the

Table 1. Transformation of vectors $\mathbf{M}$, $\mathbf{L}_\xi$ ($\xi = a, b, c$), and $\mathbf{E}$ under the action of elements of symmetry groups $Pnmm$ and $P4_2/mnm$

Γ_i	$\mathbf{M}, \mathbf{L}, \mathbf{E}$	$\bar{1}\ 2_{1x} 2_z$	4_{2z}
Γ_1	M_x, L_{cy}	$+1+1-1$	M_y, L_{cx}
Γ_2	M_y, L_{cx}	$+1-1-1$	$-M_x, -L_{cy}$
Γ_3	M_z	$+1-1+1$	M_z
Γ_4	L_{ax}, L_{by}	$-1+1+1$	L_{by}, L_{ax}
Γ_5	L_{ay}, L_{bx}, E_z	$-1-1+1$	$-L_{bx}, -L_{ay}, E_z$
Γ_6	L_{az}, E_y	$-1-1-1$	$L_{bz}, -E_x$
Γ_7	L_{bz}, E_x	$-1+1-1$	$-L_{az}, E_y$
Γ_8	L_{cz}	$+1+1+1$	$-L_{cz}$

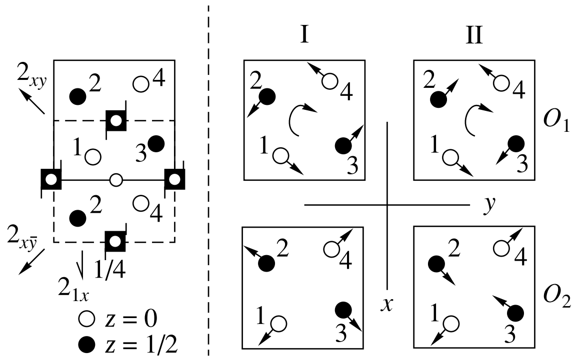


Fig. 2. Two square (or rectangular cross) exchange magnetic structures I and II forming an exchange multiplet (doublet). O_1 and O_2 are two orientations states of each of these exchange magnetic structures.

magnetoelectric and antiferroelectric interactions. Table 1 makes it possible to write invariant expressions for the energy of these interactions:

$$\begin{aligned}\Phi_{\mathrm{ME}} &= -R[E_x(\mathbf{M}\cdot\mathbf{L}_a) + E_y(\mathbf{M}\cdot\mathbf{L}_b)] \\ &- s_1(M_xE_x + M_yE_y)(L_{ax} + L_{by}) - s_2(M_xE_y + M_yE_x) \\ &\times (L_{ay} + L_{bx}) - s_3M_zE_z(L_{ax} + L_{by}),\end{aligned} \tag{6}$$

$$\begin{aligned}\Phi_{\mathrm{AFE}} &= -G[E_x(\mathbf{L}_b\cdot\mathbf{L}_c) + E_y(\mathbf{L}_a\cdot\mathbf{L}_c)] \\ &- f_1(L_{cx}E_x + L_{cy}E_y)(L_{ay} + L_{bx}) \\ &- f_2(L_{cx}E_y + L_{cy}E_x)(L_{ax} + L_{by}) \\ &- f_3L_{cz}E_z(L_{ay} + L_{bx}).\end{aligned} \tag{7}$$

Here, the terms containing coefficients R and G are of exchange origin, while the remaining terms are of relativistic origin, so that $R, G \gg s, f$.

In the spin dynamics, the magnetoelectric and antiferroelectric interactions make it possible to excite oscillations of magnetic moments by the ac electric field $\mathbf{E}(t)$. For example, if vectors $\mathbf{L}_a^0$ and/or $\mathbf{L}_b^0$ differ from zero in the ground state of an antiferromagnet, these interactions produce an ordinary magnetoelectric effect: by solving the Landau–Lifshitz equations together with expressions (6) and (7), we can determine the oscillations of magnetization $\mathbf{M}(t)$ and antiferromagnetism vectors $\mathbf{L}_\xi$ ($\xi = a, b, c$) excited by field $\mathbf{E}(t) \propto \exp(-i\omega t)$.

Thus, in the model adopted here, the compound Nd_2CuO_4 has a tetragonal crystal lattice with symmetry $P4_2/mnm$ (D_{4h}^{14}) [7–9]. After the structural transition, both paramagnetic and antiferromagnetic ($T_N = 245$ K) phases are described by the same group $P4_2/mnm$ belonging to the tetragonal system. The primitive cell containing two copper atoms in the phase $I4/mmm$ increases fourfold in the phase $P4_2/mnm$ and now contains four copper atoms [7–9].

It should be emphasized once again that magnetoelectric effects in the noncollinear antiferromagnet Nd_2CuO_4 can exist only in the presence of the above spontaneous phase transition of the crystal lattice, which occurs in this compound at $T_c = 300$ K. If, however, group (1) were preserved, antitranslations forbidding the magnetoelectric effects would appear in the magnetic symmetry group [5, 7, 9]. In addition, the magnetoelectric and antiferroelectric interactions are impossible for group (1) if only because magnetic copper ions occupy centrosymmetric positions in this case [10].

Thus, the question about the crystal symmetry group ((1) or (2)) for Nd_2CuO_4 (for copper ions) is answered by the absence or presence of the magnetoelectric effect in the antiferromagnetic phase. The authors of [11, 12] tried to find the answer to this question experimentally.

Indeed, the magnetoelectric and antiferroelectric interactions are absent for group (1) and are present for group (2). This means that the magnetoelectric and antiferroelectric effects are absent in the former case, but can take place in the latter. We will consider the results of measurements [11, 12] at a later stage proceeding from the $P4_2/mnm$ structure adopted in [7–9] and used in this article. These results are not that simple and clear and require a serious analysis of the theory together with the experiment.

Let us consider in greater detail the noncollinear antiferromagnetic structure Nd_2CuO_4 in the temperature range $1.5 < T < 245$ K [7, 9] for symmetry group (2) adopted by us. Figure 2 shows Cu^{2+} ions projected onto the xy plane in the form of light circles for $z = 0$ and dark circles for $z = c/2$. As compared to Fig. 1, the origin is displaced to point (–1/2, 0, 0). The magnetic structure of Nd_2CuO_4 depicted in Fig. 2 can be conditionally called quadratic (all $\mathbf{M}_\nu$ vectors connected successively with one another form a square) or rectangular cross (all $\mathbf{M}_\nu$ vectors with their origins brought to a single point form a rectangular cross).

Columns I and II in Fig. 2 correspond to different exchange magnetic structures, while the upper and lower lines (O_1 and O_2) correspond to different orientational states. The stability of these states is determined by the magnetoanisotropic interaction. According to the results obtained by different authors who studied microscopic models, these interactions include the dipole interactions between ions, the anisotropic exchange, and pseudodipole interactions [3, 4]. We will not consider these interactions explicitly since we are planning to construct a purely symmetric phenomenological theory.

The exchange magnetic structures I and II correspond to the same exchange energy (the angles between the magnetic moments are identical for both structures and equal to 90°), but are characterized by different magnetic symmetries. Indeed, structure I contains symmetry element 4_{2z}, while in structure II this element is replaced by $4'_{2z} = 4_{2z} \cdot 1'$. Moreover, if we introduce the chirality vector [8]

$$\begin{aligned}\mathbf{Q} &= -\frac{1}{(2M_0)^2} \\ &\times(\mathbf{M}_1\times\mathbf{M}_2+\mathbf{M}_2\times\mathbf{M}_4+\mathbf{M}_4\times\mathbf{M}_3+\mathbf{M}_3\times\mathbf{M}_1) \\ &= \frac{1}{8M_0^2}\mathbf{L}_a\times\mathbf{L}_b,\end{aligned}\tag{8}$$

these structures will have different chiralities. The second part of this equality corresponds to the equimodular model, when $|\mathbf{M}_\nu| = M_0$ ($\nu = 1, 2, 3, 4$). For the normalization of $\mathbf{Q}$ introduced here (for which $\mathbf{Q}^2 = 1$), components Q_x and $Q_y = 0$ for both structures, while $Q_z = 1$ for structure I and $Q_z = -1$ for structure II.

3. THERMODYNAMIC POTENTIAL

To define the ground state and spectrum of spin waves (magnons), we must write the thermodynamic potential Φ (we confine out analysis to density $\Phi(\mathbf{r})$ = const, which is constant in space). Apart from the magnetoelectric and antiferroelectric interactions, the potential must include the magnetic energy (exchange) and the magnetic anisotropy:

$$\begin{aligned}\Phi_{\mathrm{mag}} &= \frac{1}{2}A_M\mathbf{M}^2+\frac{1}{2}A_a(\mathbf{L}_a^2+\mathbf{L}_b^2) \\ &+\frac{1}{2}A_c\mathbf{L}_c^2+D(\mathbf{L}_a^2-\mathbf{L}_b^2)^2 \\ &+q_aQ_z+q_s(L_{ax}L_{by}+L_{ay}L_{bx}) \\ &+\frac{1}{2}K_1(L_{ax}^2+L_{by}^2)+\frac{1}{2}K_2(L_{ay}^2+L_{bx}^2) \\ &+\frac{1}{2}g(M_x^2+M_y^2)+\frac{1}{2}r(L_{cx}^2+L_{cy}^2) \\ &+p(M_xL_{cy}+M_yL_{cx}).\end{aligned}\tag{9}$$

In addition to the bilinear exchange (the terms with coefficients A), the biquadratic exchange (the term with D) is also taken into account. The term associated with chirality has coefficient q_a, and the remaining terms are the relativistic magnetic anisotropy (which contains relativistic terms of the type of quadratic interionic anisotropy), anisotropic exchange, and the pseudodipole interaction. In the presence of a magnetic field $\mathbf{H}$, the total potential Φ also includes the Zeeman energy

$$\Phi_Z = -\mathbf{M}\cdot\mathbf{H}.\tag{10}$$

Thus, the total thermodynamic potential (density) is given by

$$\Phi = \Phi_{\mathrm{mag}}+\Phi_Z+\Phi_{\mathrm{ME}}+\Phi_{\mathrm{AFE}}.\tag{11}$$

All the terms are invariant to elements $\bar{1}$, 2_{1x}, and 4_{2z} which are regarded as generators of the $P4_2/mnm$ group (see Table 1). Among other things, invariance of the chirality

$$Q_z = \frac{1}{8M_0^2}(L_{ax}L_{by}-L_{ay}L_{bx})\tag{12}$$

is also important. For structures I and II considered here, we have $Q_z = \pm 1$. On the other hand, transposition of indices a and b in expression (12) leads to sign reversal of Q_z, which corresponds to a phase transition between two chiral exchange magnetic structures I and II.

Analogously to the Dzyaloshinskii–Moriya antisymmetric exchange, chirality (12) is an exchange-relativistic quantity; consequently, magnetic interactions of various origin in energy (9) are characterized by the inequalities

$$A \gg D,\ q_a \gg q_s, K_1, K_2, g, r, p.\tag{13}$$

The equimodular model adopted above ($\mathbf{M}_\nu^2 = M_0^2$) corresponds to the conditions

$$\mathbf{M}^2+\sum_\xi \mathbf{L}_\xi^2 = (4M_0)^2,\tag{14}$$

$$\begin{aligned}\mathbf{M}\cdot\mathbf{L}_a+\mathbf{L}_b\cdot\mathbf{L}_c &= \mathbf{M}\cdot\mathbf{L}_b+\mathbf{L}_a\cdot\mathbf{L}_c \\ &= \mathbf{M}\cdot\mathbf{L}_c+\mathbf{L}_a\cdot\mathbf{L}_b = 0,\end{aligned}\tag{15}$$

as well as the Landau–Lifshitz equations for vectors $\mathbf{M}$ and $\mathbf{L}_\xi$ ($\xi = a, b, c$), which are written, for example, in [13] for antiferromagnets with four sublattices.

4. EXCHANGE MAGNETIC STRUCTURES AND ORIENTATIONAL STATES

The conditions for the realization and stability of the exchange magnetic structures and orientational states $\mathrm{I}O_1$, $\mathrm{I}O_2$ ($Q_z = 1$) and $\mathrm{II}O_1$, $\mathrm{II}O_2$ ($Q_z = -1$), which are presented in Fig. 2, were studied in [7]. We will not repeat these conditions here and will first use the above approximation, in which the role of chirality is manifested most clearly. In addition to the exchange terms, we will retain only one (semiexchange–semirelativistic) term associated with chirality Q_z in the magnetic part of thermodynamic potential (9). In accordance with inequalities (13), the remaining purely relativistic terms will be omitted at this stage, so that the potential assumes the form

$$\Phi_{\mathrm{mag}} \approx \mathrm{const} + \frac{1}{2}\tilde{A}_M \mathbf{M}^2 + \frac{1}{2}\tilde{A}_c \mathbf{L}_c^2 + D(\mathbf{L}_a^2 - \mathbf{L}_b^2)^2 + q_a Q_z. \tag{16}$$

Here, we also take into account relation (14) for the equimodular model, using this relation to eliminate the exchange terms containing A_a. We assume in this case that

$$A_a < 0, \quad A_M > 0, \quad A_c > 0, \quad D > 0 \tag{17}$$

and introduce new parameters

$$\tilde{A}_M = A_M - A_a, \quad \tilde{A}_c = A_c - A_a, \tag{18}$$

both of which are positive.

Minimizing potential (16), we obtain, among other things, two exchange magnetic structures in the easy-plane state, viz., the phases with $\mathbf{L}_a^2 = \mathbf{L}_b^2 = 8M_0^2 \equiv L^2$ for $\mathbf{L}_a \perp \mathbf{L}_b$, $\mathbf{L}_a$ and $\mathbf{L}_b \perp z$, $M = \mathbf{L}_c = 0$, which precisely correspond to two chiral structures I ($Q_z = 1$) for $q_a < 0$ and II ($Q_z = -1$) for $q_a > 0$, which are shown in Fig. 2. For the orientational state IO_1, we have $L_{ax}^0 = L_{by}^0 \equiv (\pm)L$, while for the I$O_2$ state with the same energy, we obtain $-L_{ay}^0 = L_{bx}^0 \equiv (\pm)L$. Analogously, for the state IIO_1, we have $L_{ay}^0 = L_{bx}^0 \equiv (\pm)L$, while for the state II$O_2$, we obtain $L_{ax}^0 = -L_{by}^0 \equiv (\pm)L$.

The signs (±) in the above equalities indicate that each of the four orientational states contains two possibilities differing only in the sign of the vectors corresponding to the ground state. This points to the existence of domains for which the magnetization associated with the magnetoelectric effect also differs only in sign. If the number of such domains (say, with vectors $L_{ax}^0 = L_{by}^0$) in the orientational state under investigation is the same (the IO_1 structure), the total magnetization is zero (see below). To detect the linear magnetoelectric effect, poling of the sample must be first carried out; this can be done, for example, by cooling the sample from the paramagnetic region in an appropriate external magnetic field (thermomagnetic treatment). Apparently, this was not done in [11, 12].

Naturally, it should be borne in mind that, strictly speaking, it is the relativistic anisotropy (which is ignored in expression (16)) that determines as to which of the two orientational states, O_1 or O_2, is realized for each exchange magnetic structure.

In accordance with Table 1, structure IO_1 (from the position of the rhombic group *Pnmm*) corresponds to representation $\Gamma_4(L_{ax}^0, L_{by}^0)$; although this representation becomes two-dimensional upon a transition to the tetragonal group $P4_2/mnm$ after the addition of the 4_{2z} axis, it contains the same nonzero components of vectors $\mathbf{L}_a^0$ and $\mathbf{L}_b^0$. The latter statement means that the magnetic group for this structure (we will call it the $\Gamma_4(L_{ax}^0, L_{by}^0)$ phase) is $\bar{1}'4_{2z}2_{1x}$. The IIO_1 structure corresponds to the representation $\Gamma_5(L_{ay}^0, L_{bx}^0)$, in which the 4_{2z} axis transforms the vector components constituting it into one another, but with opposite signs. The magnetic group of the corresponding phase is $\bar{1}'4'_{2z}2'_{1x}$.

The other two phases (which can be written as $\Gamma_4(-L_{ay}^0, L_{bx}^0)$ and $\Gamma_5(L_{ax}^0, -L_{by}^0)$ in the orientational state O_2) can be obtained from the above phases by rotating all magnetic moments through 90°; henceforth, we will consider only the phases in the orientational state O_1 (i.e., $\Gamma_4(-L_{ax}^0, L_{by}^0)$ and $\Gamma_5(L_{ay}^0, -L_{bx}^0)$ with chirality $Q_z = 1$ and $Q_z = -1$, respectively).

It should be noted that the semiexchange–semirelativistic interaction with the chirality energy $\Phi_Q = q_a Q_z$ is an analog of the pseudodipole interaction with Hamiltonian H_Q introduced in [4] for explaining the noncollinearity of the magnetic structure of Nd_2CuO_4. This can easily be verified if, taking into account relations (12) and (3), we write energy Φ_Q in the form

$$\Phi_Q = -\frac{q_a}{M_0^2}(M_{1x}^0 M_{2y}^0 - M_{1y}^0 M_{2x}^0) = q_a \sin(\varphi_1 - \varphi_2),$$

where φ_1 and φ_2 are the azimuth angles of vectors $\mathbf{M}_1^0 = -\mathbf{M}_4^0$ and $\mathbf{M}_2^0 = -\mathbf{M}_3^0$. Consequently, energy Φ_Q, as well as H_Q, describes the interaction of the magnetic moments of the Cu^{2+} ions lying in the adjacent planes $z = 0$ and $z = c/2$ (see Fig. 2, states O_1). The minimum value of energy Φ_Q corresponds to $\varphi_2 - \varphi_1 = -\pi/2$ for $q_a < 0$ (phase Γ_4, $Q_z = 1$) or $\varphi_2 - \varphi_1 = \pi/2$ for $q_a > 0$ (phase Γ_5, $Q_z = -1$), corresponding to the noncollinear magnetic structures shown in Fig. 2 (states O_1).

5. SPIN-WAVE REPRESENTATIONS (OSCILLATION MODES): FREQUENCIES AND SUSCEPTIBILITIES TO FIELD $\mathbf{E} = \mathbf{E}(t)$

In calculating the natural oscillation frequencies, it is convenient to use the concept of spin-wave representations [9, 13–16], which makes it possible to separate vibrational variables into independent groups corresponding to vibrational modes even before the equations of motion are written. In the general case, the group-theoretical approach is used for this purpose, which is based on the theory of corepresentations of magnetic groups [14, 15]. However, for the rhombic symmetry, the following simple algorithm was proposed [9, 13], in which Table 1 (the part corresponding to the rhombic group) is used. The same spin-wave representation of phase Γ_n under investigation corresponds to variables from two lines of the table (Γ_m and $\Gamma_{m'}$) for which the product of numbers (+1 or –1) gives the numbers of line Γ_n. This can be written schematically as

$$\Gamma_m \cdot \Gamma_{m'} = \Gamma_n. \tag{19}$$

The case when $m = m'$ is also possible.

Unfortunately, an analogous simplified rule for determining the spin-wave representations for tetragonal crystals has not been formulated as yet. It will be shown below, however, that even the spin-wave representations determined for the corresponding rhombic subgroup provide information sufficient for obtaining representations for the tetragonal case by using the specific form of the thermodynamic potential and the corresponding symmetry of the problem.

Let us demonstrate this for structures (phases) $\Gamma_4(L^0_{ax}, L^0_{by})$ and $\Gamma_5(L^0_{ay}, L^0_{bx})$.

5.1. Spin Oscillations in the $\Gamma_4(L^0_{ax}, L^0_{by})$ Phase Induced by Field $\mathbf{E}(t)$

In accordance with the rule formulated above for the rhombic subgroup, Table 1 for this phase gives the following four spin-wave representations (modes): antimagnons[1]

$$\Gamma_{48}(\Delta L_{ax}, \Delta L_{by}, \Delta L_{cz}) \tag{20}$$

and three quasi-antiferromagnon modes

$$\Gamma_{17}(\Delta M_x, \Delta L_{cy}, \Delta L_{bz}), \tag{21}$$

$$\Gamma_{26}(\Delta M_y, \Delta L_{cx}, \Delta L_{az}), \tag{22}$$

$$\Gamma_{35}(\Delta M_z, \Delta L_{ay}, \Delta L_{bx}), \tag{23}$$

each of which has one component of vector $\mathbf{M}$ among vibrational variables.

What will happen to these modes after the inclusion of the 4_{2z} axis (i.e., in the case of the tetragonal group)? To be more precise, how are these modes transformed under the action of the 4_{2z} symmetry axis (since the exact symmetry of the $\Gamma_4(L^0_{ax}, L^0_{by})$ phase is determined by the magnetic group $\bar{1}'4_{2z}2_{1x}$)? The answer to these questions is quite favorable: each of the modes Γ_{48} and Γ_{35} is transformed into itself, while modes Γ_{17} and Γ_{26} are transformed into each other! Thus, modes (20)–(23) remain unchanged for the tetragonal phase also and only modes Γ_{17} and Γ_{26} are transformed into each other under the action of element 4_{2z}.

To solve the problem of natural oscillations excited by field $\mathbf{E}(t) \propto \exp(-i\omega t)$ in the $\Gamma_4(L^0_{ax}, L^0_{by})$ phase, we must single out the terms in thermodynamic potential (9), which are quadratic in the vibrational variables appearing in modes (20)–(23) of this phase and supplement the corresponding terms from expressions (6) and (7) for Φ_{ME} and Φ_{AFE}, respectively.

Thus, we find that

$$\begin{aligned}\Phi_2(\pm) = {}& \frac{1}{2}[\tilde{A}_M \Delta M_z^2 + \tilde{A}(\Delta L_{cx}^2 + \Delta L_{cy}^2) \\ & + \tilde{A}_c \Delta L_{cz}^2 + D^+ \Delta L_{by}^2 + K^+ \Delta L_{bx}^2 \\ & + q^+(\Delta L_{az}^2 + \Delta L_{bz}^2)] - (\tilde{G} - \tilde{R}) \\ & \times (E_x \Delta L_{cy} + E_y \Delta L_{cx}) - 2\tilde{s}_3 \Delta M_z E_z.\end{aligned} \tag{24}$$

In the derivation of this equation, relations (14) and (15) of the equimodular model play a significant role. Owing to these relations, variable ΔM_x, ΔM_y, ΔL_{ax}, and ΔL_{ay} are eliminated and an exchange contribution from Φ_{ME} (term $\tilde{R}$ in Eq. (24)) appears. In addition to parameters $\tilde{A}_M$ and $\tilde{A}_c$ (18), one more exchange parameter, viz., their sum

$$\tilde{A} = \tilde{A}_M + \tilde{A}_c, \tag{25}$$

also appears. The remaining notation in (24) is as follows:

$$\begin{aligned} D^+ &= 4[8DL^2 - (\tilde{q}_a + q_s)], \\ K^+ &= 2(K_2 - K_1 - 2q_s), \\ q^+ &= -(\tilde{q}_a + q_s + K_1), \\ \tilde{G} - \tilde{R} &= (\pm)(G - R)L, \quad \tilde{s}_3 = (\pm)s_3 L, \end{aligned} \tag{26}$$

$\tilde{q}_a = q_a/L^2$, $\tilde{q}_a + q_s < 0$, $K^+ > 0$, and $K_1 < 0$ (see stability conditions (63) and (64) below for phase Γ_4). Here, the (±) signs indicate allowance for domains with positive ($+\mathbf{L}^0$) and negative ($-\mathbf{L}^0$) values of the vectors of the ground state, while the plus and minus signs (separately) in the argument as well as in the superscripts of physical quantities indicate that they correspond to phase Γ_4 or Γ_5 with chirality $Q_z = +1$ or $Q_z = -1$, respectively.

Using expression (24) for Φ_2, we can now find solutions to the Landau–Lifshitz equations [13] for all four modes (20)–(23).

The antimagnon mode $\Gamma_{48}(\Delta L_{ax}, \Delta L_{by}, \Delta L_{cz})$ in the $\Gamma_4(L^0_{ax}, L^0_{by})$ phase is not excited by electric field $\mathbf{E}(t)$

---

[1] It was agreed earlier [13, 17] that this term is applied to the mode for which vector $\mathbf{M}$ does not participate in oscillations and only vectors $\mathbf{L}$ oscillate.

since expressions (6) and (7) do not contain any terms linear in its variables.

For the $\Gamma_{17}(\Delta M_x, \Delta L_{cy}, \Delta L_{bz})$ mode excited by field **E** parallel to the x axis, we obtain

$$\begin{gathered}\Delta M_x = \alpha_{xx}E_x = -\frac{(\tilde{G}-\tilde{R})}{\tilde{A}}\frac{\omega_{\mathrm{AQ}}^2(+)}{\omega_{\mathrm{AQ}}^2(+)-\omega^2}E_x, \\ \Delta L_{bz} = -i\frac{\omega}{\omega_0 q^+}\Delta M_x, \quad \Delta L_{cy} = -\Delta M_x,\end{gathered} \tag{27}$$

where α_{xx} is the corresponding component of the magnetoelectric susceptibility tensor. If we take into account the presence of the above-mentioned domains, the total magnetization in compliance with relations (27) will be zero in the case of complete compensation of domains (see above): $\sum_{(\pm)}\Delta M_x = 0$. However, incomplete compensation (a structure with predominance of domains with identical sign of vectors $\mathbf{L}_{a,\,b}$ over other domains) is also possible; in this case, $\sum_{(\pm)}\Delta M_x \neq 0$.

In view of the above-mentioned relation between modes $\Gamma_{17}(\Delta M_x, \Delta L_{cy}, \Delta L_{bz})$ and $\Gamma_{26}(\Delta M_y, \Delta L_{cx}, \Delta L_{az})$, we can immediately write the corresponding expressions for the latter mode. We will not do it here, bearing in mind that the Γ_{26} mode can be obtained from the Γ_{17} mode by the symmetry operation 4_{2z} (see Table 1); it should only be noted that these modes are degenerate, i.e., having the same resonance frequency

$$\omega_{\mathrm{AQ}} = \omega_0\sqrt{\tilde{A}q^+} \equiv \omega_{\mathrm{AQ}}(+), \tag{28}$$

where $\omega_0 = \gamma L$, γ being the absolute value of the magnetomechanical ratio.

It should be noted that the magnetoelectric susceptibility α_{xx} in formula (27) is determined by the exchange part of the magneto- and antiferroelectric interactions (the terms containing $\tilde{G} - \tilde{R}$ in formula (24)); consequently, we can hope that the value of this susceptibility (in the absence of domains or in the presence of a grain structure) is much higher than the values of susceptibilities determined by relativistic interactions.

Finally, for the fourth (quasi-antiferromagnon) mode $\Gamma_{35}(\Delta M_z, \Delta L_{ay}, \Delta L_{bx})$, the Landau–Lifshitz equations [13], taking into account (24), give

$$\begin{gathered}\Delta M_z = \alpha_{zz}E_z = \frac{2\tilde{s}_3}{\tilde{A}_M}\frac{\omega_{\mathrm{AF}}^2(+)}{\omega_{\mathrm{AF}}^2(+)-\omega^2}E_z, \\ \Delta L_{ay} = -\Delta L_{bx} = -\frac{i\omega}{\omega_0 K^+}\Delta M_z.\end{gathered} \tag{29}$$

This mode is excited by field **E** parallel to the z axis due to the relativistic magnetoelectric interaction (the term containing $\tilde{s}_3$ in (24)). Its frequency

$$\omega_{\mathrm{AF}}(+) = \omega_0\sqrt{\tilde{A}_M K^+} \tag{30}$$

is determined by the geometrical mean of the exchange constant $\tilde{A}_M$ from formula (18) and the linear combination K^+ from relation (26), which consists of relativistic anisotropy constants.

5.2. Spin Oscillations Excited in the $\Gamma_5(L_{ay}^0,\, L_{bx}^0)$ Phase by Field **E***(t)*

In this phase, $L_{ay}^0 = L_{bx}^0 \equiv (\pm)L$ are the nonzero components of the basis vectors of the ground state, and the magnetic subgroup of this phase is $\bar{1}'4'_{2z}2'_{1x}$. As in the case of the Γ_4 mode, we find from Table 1 the spin-wave representations

$$\Gamma_{58}(\Delta L_{ay}, \Delta L_{bx}, \Delta L_{cz}) \tag{31}$$

for the antimagnon mode,

$$\begin{gathered}\Gamma_{16}(\Delta M_x, \Delta L_{cy}, \Delta L_{az}), \\ \Gamma_{27}(\Delta M_y, \Delta L_{cx}, \Delta L_{bz})\end{gathered} \tag{32}$$

for the quasi-antiferromagnon modes connected with each other via the symmetry operation $4'_{2z}$, and

$$\Gamma_{34}(\Delta M_z, \Delta L_{ax}, \Delta L_{by}) \tag{33}$$

for one more quasi-antiferromagnon mode.

The quadratic thermodynamic potential determined from formula (11), taking into account relations (14) and (15), can be represented in terms of vibrational variables of these modes in the form

$$\begin{gathered}\Phi_2(\pm) = \frac{1}{2}[\tilde{A}_M\Delta M_z^2 + \tilde{A}(\Delta L_{cx}^2 + \Delta L_{cy}^2) \\ + \tilde{A}_c\Delta L_{cz}^2 + D^-\Delta L_{bx}^2 + K^-\Delta L_{by}^2 \\ + q^-(\Delta L_{az}^2 + \Delta L_{bz}^2)] \\ -(\tilde{G}-\tilde{R})(E_x\Delta L_{cx} + E_y\Delta L_{cy}) - 2\tilde{f}_3\Delta L_{cz}E_z.\end{gathered} \tag{34}$$

Here, we used the notation

$$\begin{gathered}D^- = 4(8DL^2 + \tilde{q}_a - q_s), \\ K^- = 2(K_1 - K_2 - 2q_s), \\ q^- = \tilde{q}_a - q_s - K_2, \quad \tilde{f}_3 = (\pm)f_3L,\end{gathered} \tag{35}$$

where $\tilde{q}_a - q_s > 0$, $K^- > 0$, and $K_2 < 0$ (see the stability conditions (65), (66) for the Γ_5 phase).

Solving the Landau–Lifshitz equations [13] successively for the Γ_{58} modes from relation (31), Γ_{16} and Γ_{27} modes from relation (32), and the Γ_{34} mode from relation (33), we arrive at the following results.

The antimagnon mode $\Gamma_{58}(\Delta L_{ay}, \Delta L_{bx}, \Delta L_{cz})$ containing oscillations ΔL_{cz} is excited by field $\mathbf{E}(t)$ directed along the z axis (expression (34), the term containing coefficient $\tilde{f}_3$). Unfortunately, this is a relativistic term. This gives

$$\begin{gathered} \Delta L_{cz} = \beta_{zz} E_z = 2\frac{\tilde{f}_3}{\tilde{A}_c} \frac{\omega_L^2(-)}{\omega_L^2(-) - \omega^2} E_z, \\ \Delta L_{bx} = i\frac{\omega}{\omega_0 D^-}\Delta L_{cz}, \quad \Delta L_{ay} = -\Delta L_{bx}, \end{gathered} \tag{36}$$

where β_{zz} is the component of the antimagnon-electric susceptibility tensor and

$$\omega_L(-) = \omega_0\sqrt{\tilde{A}_c D^-} \tag{37}$$

is the antimagnon resonance frequency. The second factor in the radicand (D^- from expressions (35)) contains, apart from biquadratic exchange D the chiral term with $\tilde{q}_a$. These two terms may be comparable in value, but they are usually smaller than the bilinear exchange. In this respect, the magnet under consideration with an exchange-noncollinear magnetic structure differs from collinear or weakly relativistically noncollinear structure considered earlier [13, 17]. In accordance with adopted inequalities (13), the antimagnon frequency ω_L (37) may turn out to be much smaller than in those cases. It should be noted that the antimagnon mode $\Gamma_{48}(\Delta L_{ax}, \Delta L_{by}, \Delta L_{cz})$, which is not excited by the electric field, $\mathbf{E}(t)$, has natural frequency $\omega_L(+)$ similar to frequency $\omega_L(-)$. It can be determined from formula (37) by replacing D^- (35) in this formula by D^+ (26) (see formula (47) below).

It should be noted that the antiferromagnetism vector $\Delta\mathbf{L}$ cannot be a directly observable quantity in antimagnon resonance. In this respect, it is more convenient to use the effective polarization vector (in the present case, its component

$$P_z = -\frac{\partial\Phi_2}{\partial E_z} = 2\tilde{f}_3\Delta L_{cz} = \kappa_{zz}E_z, \tag{38}$$

where $\kappa_{zz} = 2\tilde{f}_3\beta_{zz}$ is the dielectric susceptibility tensor component). In this case, we can calculate, for example, the absorbed power

$$Q = -\overline{P_z\dot{E}_z} \tag{39}$$

(the bar indicates averaging over time $t \gg 2\pi/\omega$). Naturally, we must first take into account damping in formula (39). In the simplest form (in Bloch's sense), this can be reduced to the substitution $\omega \longrightarrow \omega + i\Gamma$ in formula (39), where Γ is the resonance half-width. As a result, we obtain from Eq. (39)

$$Q = \omega\tilde{f}_3\beta''_{zz}|E_z|^2, \tag{40}$$

where β''_{zz} is the imaginary part of β_{zz} from formula (36) (taking into account damping). The above arguments should be taken into account in experiments on observing the antimagnon resonance since, in contrast to the antiferromagnetic resonance, resonance for electric polarization and not for magnetization takes place in field $\mathbf{E}(t)$ in this case.

Modes Γ_{16} and Γ_{27} (32) transformed into each other under transformation $4'_{2z}$, have natural frequency $\omega_{\mathrm{AQ}}(-)$, which can be obtained from expression (28) for $\omega_{\mathrm{AQ}}(+)$ in the case when Γ_{17} is defined by formula (21) and Γ_{26} is defined by formula (22) if we replace q^+ by q^- in the latter formula. In particular, for the mode $\Gamma_{16}(\Delta M_x, \Delta L_{cy}, \Delta L_{az})$ excited by field $\mathbf{E}$ directed along the y axis, we have

$$\begin{gathered} \Delta M_x = \alpha_{xy}E_y = -\frac{\tilde{G} - \tilde{R}}{\tilde{A}}\frac{\omega_{\mathrm{AQ}}^2(-)}{\omega_{\mathrm{AQ}}^2(-) - \omega^2}E_y, \\ \Delta L_{cy} = -\Delta M_x, \\ \Delta L_{az} = -i\frac{\omega}{\omega_0 q^-}\Delta M_x, \quad \omega_{\mathrm{AQ}}(-) = \omega_0\sqrt{\tilde{A}q^-}. \end{gathered} \tag{41}$$

The corresponding expressions for the mode $\Gamma_{27}(\Delta M_y, \Delta L_{cx}, \Delta L_{bz})$ can be obtained from these formulas by the action of the symmetry operation $4'_{2z}$ (see Table 1).

Let us now consider the mode $\Gamma_{34}(\Delta M_z, \Delta L_{ax}, \Delta L_{by})$. Its variables ($L_{ay}^0 = L_{bx}^0 \equiv (\pm)L$ in the ground state) do not appear in expressions (6) for Φ_{ME} and (7) for Φ_{AFE}; consequently, it is not excited by field $\mathbf{E}(t)$. The frequency $\omega_{\mathrm{AF}}(-)$ of this mode, as well as the frequency $\omega_{\mathrm{AF}}(+)$ of the Γ_{35} mode (see relation (30)), is determined by the relativistic constants enhanced by exchange $\tilde{A}_M$ (see formula (60) below).

6. EXCITATION OF MAGNETIZATION BY AC MAGNETIC FIELD $\mathbf{H}(t)$ UPON APPLICATION OF CONSTANT ELECTRIC FIELD $\mathbf{E}^0$

Although the conventional linear magnetoelectric effect (in field $\mathbf{E}(t)$) could not be reliably observed in [11, 12], another magnetoelectric effect quadratic in

$\mathbf{E}(t)$ was observed for sure. The theory predicts this effect from the following considerations.

Let us define the varying magnetization $\mathbf{M}(t)$,

$$\mathbf{M} = -\frac{\partial\Phi}{\partial\mathbf{h}}, \qquad (42)$$

emerging due to the Zeeman and magnetoelectric interactions when a varying magnetic field $\mathbf{H} = \mathbf{h}(t)$ and a constant electric field $\mathbf{E} = \mathbf{E}^0 = \text{const}$ are applied simultaneously. The energy of the magnetoelectric interaction in this case can be written in the form

$$\Phi'_{\text{ME}} = -R'[E_x^0(\mathbf{h}\cdot\mathbf{L}_a) + E_y^0(\mathbf{h}\cdot\mathbf{L}_b)]. \qquad (43)$$

Here, the meaning and magnitude of coefficient R' differ from those of the corresponding coefficient R in formula (6) and relation (43) is analogous to the terms in formula (6) containing R only in symmetry. In expression (43), the relativistic terms analogous to those from (6) after the substitution of $\mathbf{h}$ for $\mathbf{M}$ are not included since these terms make zero contribution to quantities ΔM_i, $\Delta L_{\xi i}$ ($\xi = a, b, c$), and $\mathbf{M}(t)$ in the equimodular model. Taking into account relations (10) and (43), we obtain from Eq. (42) the expression for the total magnetization,

$$M_i(t) = \Delta M_i + R'(E_x^0\Delta L_{ai} + E_y^0\Delta L_{bi}), \qquad (44)$$

in which the second term, including the entire expression in the parentheses, describes the contribution of oscillations of antiferromagnetism vectors $\Delta L_{\xi i}$ to quantity $\mathbf{M}(t)$. This contribution is associated with the exchange part of the magnetoelectric interaction with constant R'. Here, ΔM_i, $\Delta L_{\xi i}$ ($\xi = a, b$) are the vibrational variables corresponding to modes (20)–(23) in the $\Gamma_4(L_{ax}^0, L_{by}^0)$ phase and to modes (31)–(33) in the $\Gamma_5(L_{ay}^0, L_{bx}^0)$ phase. We will find these variables from the Landau–Lifshitz equations [13], using the thermodynamic potential Φ composed from energies Φ_{mag} (9), Φ_Z (10), and Φ_{ME} (43). Here, we disregard the antiferroelectric interaction energy Φ_{AFE} (7) since we ignore a weak effect of the constant electric field on the ground state and frequency of oscillations emerging when the term Φ_{AFE} is taken into account in the thermodynamic potential Φ. It can be proved that this effect is relatively weak indeed.

6.1. $\Gamma_4(L_{ax}^0, L_{by}^0)$ *Phase*

Let us first consider the antimagnon mode $\Gamma_{48}(\Delta L_{ax}, \Delta L_{by}, \Delta L_{cz})$. The Landau–Lifshitz equations [13] taking into account expressions (9)–(11) and (43) in this case have the form

$$\Delta L_{ax} = -\Delta L_{by} = \frac{\omega_0^2 R'\tilde{A}_c(E_x^0 h_x - E_y^0 h_y)}{\omega_L^2(+) - \omega^2}, \qquad (45)$$

$$\Delta L_{cz} = -\frac{i\omega}{\omega_0\tilde{A}_c}\Delta L_{ax}, \qquad (46)$$

where

$$\omega_L(+) = \omega_0\sqrt{\tilde{A}_c D^+} \qquad (47)$$

is the antimagnon resonance frequency for the Γ_{48} mode.

Analogously, for the $\Gamma_{17}(\Delta M_x, \Delta L_{cy}, \Delta L_{bz})$ mode, we can write

$$\Delta M_x = -\Delta L_{cy} = \frac{i\omega\omega_0 R' E_y^0 h_z + \omega_0^2 q^+ h_x}{\omega_{\text{AQ}}^2(+) - \omega^2}, \qquad (48)$$

$$\Delta L_{bz} = \frac{\omega_0^2\tilde{A}R' E_y^0 h_z - i\omega\omega_0 h_x}{\omega_{\text{AQ}}^2(+) - \omega^2}. \qquad (49)$$

Variables for the $\Gamma_{26}(\Delta M_y, \Delta L_{cx}, \Delta L_{az})$ mode, which is degenerate with the Γ_{17} mode, are determined from expressions (48), (49) by the action of the symmetry element 4_{2z} as before (see Table 1)

Using the Landau–Lifshitz equations [13], we obtain the following expressions for the $\Gamma_{35}(\Delta M_z, \Delta L_{ay}, \Delta L_{bx})$ mode:

$$\Delta M_z = \frac{i\omega\omega_0 R'(E_x^0 h_y - E_y^0 h_x) + \omega_0^2 K^+ h_z}{\omega_{\text{AF}}^2(+) - \omega^2}, \qquad (50)$$

$$\begin{aligned}\Delta L_{ay} &= -\Delta L_{bx}\\ &= \frac{\omega_0^2\tilde{A}_M R'(E_x^0 h_y - E_y^0 h_x) - i\omega\omega_0 h_z}{\omega_{\text{AF}}^2(+) - \omega^2}.\end{aligned} \qquad (51)$$

In formulas (48)–(51), quantities $\omega_{\text{AQ}}(+)$ and $\omega_{\text{AF}}(+)$ are the natural frequencies of the corresponding modes given by formulas (28) and (30).

It follows from expressions (45), (46), and (48)–(51) that the antimagnon mode $\Gamma_{48}(\Delta L_{ax}, \Delta L_{by}, \Delta L_{cz})$ (which is not excited by the varying electric field $\mathbf{E}(t)$) is excited by a varying magnetic field $\mathbf{h}(t) \perp z$ in the presence of field $\mathbf{E}^0 \perp z$ owing to the magnetoelectric interaction. Other (quasi-antiferromagnon) modes can be excited due to both the magnetoelectric interaction (first terms in Eqs. (48)–(51)) and the Zeeman interaction (last terms). For example, the quasi-antiferromagnon mode Γ_{17} (21) is excited by magnetic field $\mathbf{h}(t) \parallel x$

Table 2. Modes of spin oscillations and electric field **E** and magnetic field **h** exciting them in the Γ_4 and Γ_5 phases. Here, $D^{\pm} = 4(8DL^2 \mp \tilde{q}_a - q_s)$, $K^{\pm} = 2(\mp K_1 \pm K_2 - 2q_s)$, $q^+ = -(\tilde{q}_a + q_s + K_1)$, and $q^- = \tilde{q}_a - q_s + K_2$

Phase stability regions	Vibrational modes	**E**	**E**^{0}, **h**	**h**	Natural frequencies
$\Gamma_4(L^0_{ax}, L^0_{by})$	$\Gamma_{48}(\Delta L_{ax}, \Delta L_{by}, \Delta L_{cz})$	–	$E^0_x h_x$, $E^0_y h_y$	–	$\omega^+_L = \omega_0\sqrt{\tilde{A}_c D^+}$
$\tilde{q}_a + q_s < 0$	$\Gamma_{17}(\Delta M_x, \Delta L_{cy}, \Delta L_{bz})$	E_x	$E^0_y h_z$	h_x	
$K^+ > 0$					$\omega^+_{\mathrm{AQ}} = \omega_0\sqrt{\tilde{A}q^+}$
	$\Gamma_{26}(\Delta M_y, \Delta L_{cx}, \Delta L_{az})$	E_y	$E^0_x h_z$	h_y	
	$\Gamma_{35}(\Delta M_z, \Delta L_{ay}, \Delta L_{bx})$	E_z	$E^0_x h_y$, $E^0_y h_x$	h_z	$\omega^+_{\mathrm{AF}} = \omega_0\sqrt{\tilde{A}_M K^+}$
$\Gamma_5(L^0_{ay}, L^0_{bx})$	$\Gamma_{58}(\Delta L_{ay}, \Delta L_{bx}, \Delta L_{cz})$	E_z	$E^0_x h_y$, $E^0_y h_x$	–	$\omega^-_L = \omega_0\sqrt{\tilde{A}_c D^-}$
$\tilde{q}_a - q_s < 0$	$\Gamma_{16}(\Delta M_x, \Delta L_{cy}, \Delta L_{az})$	E_y	$E^0_x h_z$	h_x	
$K^- > 0$					$\omega^-_{\mathrm{AQ}} = \omega_0\sqrt{\tilde{A}q^-}$
	$\Gamma_{27}(\Delta M_y, \Delta L_{cx}, \Delta L_{bz})$	E_x	$E^0_y h_z$	h_y	
	$\Gamma_{34}(\Delta M_z, \Delta L_{ax}, \Delta L_{by})$	–	$E^0_x h_x$, $E^0_y h_y$	h_z	$\omega^-_{\mathrm{AF}} = \omega_0\sqrt{\tilde{A}_M K^-}$

(owing to the Zeeman interaction) or under the simultaneous action of fields $\mathbf{h}(t) \parallel z$ and $\mathbf{E}^0 \parallel y$ (owing to the magnetoelectric interaction), while the quasi-antiferromagnon mode Γ_{35} (23) is excited by field $\mathbf{h}(t) \parallel z$ or by fields $\mathbf{h}(t) \perp z$ and $\mathbf{E}^0 \perp z$ acting together (see Table 2).

Using expressions (45), (46), and (48)–(51) for vibrational variables ΔM and ΔL_ξ ($\xi = a, b$) in expression (44), we find that the total magnetization $\mathbf{M}(t)$ is defined as

$$M_i(t) = (\chi_{ik} + \Delta\chi_{ik})h_k, \tag{52}$$

where χ_{ik} is the magnetic susceptibility tensor with diagonal components (other components are equal to zero),

$$\begin{aligned} \chi_{xx} &= \chi_{yy} = \frac{1}{\tilde{A}}\frac{\omega^2_{\mathrm{AQ}}(+)}{\omega^2_{\mathrm{AQ}}(+) - \omega^2}, \\ \chi_{zz} &= \frac{1}{\tilde{A}_M}\frac{\omega^2_{\mathrm{AF}}(+)}{\omega^2_{\mathrm{AF}}(+) - \omega^2} \end{aligned} \tag{53}$$

(susceptibilities χ_{ik} describe excitation of magnetization $\mathbf{M}(t)$ by magnetic field $\mathbf{h}(t)$ due to the Zeeman interaction); $\Delta\chi_{ik}$ is a tensor describing the magnetoelectric correction to the magnetic susceptibility with components

$$\begin{aligned} \Delta\chi_{xx} &= \frac{R'^2}{D^+}\frac{\omega^2_L(+)}{\omega^2_L(+) - \omega^2}(E^0_x)^2 \\ &+ \frac{R'^2}{K^+}\frac{\omega^2_{\mathrm{AF}}(+)}{\omega^2_{\mathrm{AF}}(+) - \omega^2}(E^0_y)^2, \end{aligned} \tag{54}$$

$$\begin{aligned} \Delta\chi_{yy} &= \frac{R'^2}{D^+}\frac{\omega^2_L(+)}{\omega^2_L(+) - \omega^2}(E^0_y)^2 \\ &+ \frac{R'^2}{K^+}\frac{\omega^2_{\mathrm{AF}}(+)}{\omega^2_{\mathrm{AF}}(+) - \omega^2}(E^0_x)^2, \end{aligned} \tag{55}$$

$$\Delta\chi_{zz} = \frac{R'^2}{q^+}\frac{\omega^2_{\mathrm{AQ}}(+)}{\omega^2_{\mathrm{AQ}}(+) - \omega^2}[(E^0_x)^2 + (E^0_y)^2], \tag{56}$$

$$\begin{aligned} \Delta\chi_{xy} = \Delta\chi_{yx} &= -\left(\frac{R'^2}{D^+}\frac{\omega^2_L(+)}{\omega^2_L(+) - \omega^2}\right. \\ &\left. + \frac{R'^2}{K^+}\frac{\omega^2_{\mathrm{AF}}(+)}{\omega^2_{\mathrm{AF}}(+) - \omega^2}\right)E^0_x E^0_y, \end{aligned} \tag{57}$$

$$\Delta\chi_{xz} = -\Delta\chi_{zx} = i\omega\omega_0 R' E_y^0 \{[\omega_{\rm AQ}^2(+) - \omega^2]^{-1} + [\omega_{\rm AF}^2(+) - \omega^2]^{-1}\}, \quad (58)$$

$$\Delta\chi_{yz} = -\Delta\chi_{zy} = -i\omega\omega_0 R' E_x^0 \{[\omega_{\rm AQ}^2(+) - \omega^2]^{-1} + [\omega_{\rm AF}^2(+) - \omega^2]^{-1}\}. \quad (59)$$

It should be noted that components χ_{ii} (54)–(56) and $\Delta\chi_{xy}$, $\Delta\chi_{yx}$ (57) describe the excitation of magnetization $M_i(t)$ by oscillations $\Delta\mathbf{L}_\xi$ ($\xi = a, b$) induced by the magnetic field $\mathbf{h}(t)$ due to direct exchange magnetoelectric interaction. The remaining components $\Delta\chi_{ik}$ (58), (59) describe the excitation of magnetization $\mathbf{M}(t)$ by magnetic field $\mathbf{h}(t)$ due to both the direct and indirect magnetoelectric interactions. For example, the first terms in expressions (58) for $\Delta\chi_{xz}$ and (59) for $\Delta\chi_{yz}$ are responsible for direct excitation of oscillations of $M_x = \Delta M_x$ (the first term in Eq. (48)) and $M_y = \Delta M_y$, belonging to modes Γ_{17} and Γ_{26}, respectively, while the second terms are responsible for the indirect excitation of magnetizations M_x and M_y by oscillations ΔL_{ay} and ΔL_{bx} of the Γ_{35} mode (the last term in Eq. (51)), which are induced not by the magnetoelectric, but by the Zeeman interaction. Analogously, the first terms in expressions (58) for $\Delta\chi_{zx}$ and (59) for $\Delta\chi_{zy}$ describe indirect magnetoelectric excitation of magnetization M_z by oscillations ΔL_{bz} of the Γ_{17} mode (the last term in Eq. (49)) and oscillations ΔL_{az} of the Γ_{26} mode caused by the Zeeman interaction, while the second term are responsible for direct magnetoelectric excitation of oscillations of $M_z = \Delta M_z$ (terms with R' in Eq. (50)) belonging to the Γ_{35} mode.

It should be noted that formulas (53)–(59) remain valid for an electric field containing, apart from constant component $\mathbf{E}^0$, a varying (quasi-static) component $\mathbf{E}(t)$ if we carry out the substitutions $\mathbf{E}^0 \longrightarrow \mathbf{E}^0 + \mathbf{E}(t)$ and $\omega \longrightarrow \omega_E + \omega_h \approx \omega_h$, where ω_E and ω_h ($\omega_E \ll \omega_h$) are the frequencies of oscillations of the electric and magnetic fields.

6.2. $\Gamma_5(L_{ay}^0, L_{bx}^0)$ Phase

Let us now consider the $\Gamma_5(L_{ay}^0, L_{bx}^0)$ phase. The antimagnon mode $\Gamma_{58}(\Delta L_{ay}, \Delta L_{bx}, \Delta L_{cz})$ of this phase, as well as the antimagnon mode $\Gamma_{48}(\Delta L_{ax}, \Delta L_{by}, \Delta L_{cz})$ of phase $\Gamma_4(L_{ax}^0, L_{by}^0)$, is excited by magnetic field $\mathbf{h}(t) \perp z$ for $\mathbf{E}^0 \perp z$ (see Table 2). The vibrational variables of this mode are determined from expressions (45) and (46) using the substitutions $\Delta L_{ax} \longrightarrow \Delta L_{ay}$, $\Delta L_{by} \longrightarrow \Delta L_{bx}$, $h_x \longleftrightarrow h_y$, $\omega \longrightarrow -\omega$, $\omega_L(+) \longrightarrow \omega_L(-)$, where $\omega_L(-)$ is the antimagnon resonance frequency for the Γ_{58} mode (see formula (37)).

The quasi-antiferromagnon mode $\Gamma_{16}(\Delta M_x, \Delta L_{cy}, \Delta L_{az})$ is excited by magnetic field $\mathbf{h}(t) \parallel x$ or by the simultaneous action of fields $\mathbf{E}^0 \parallel x$ and $\mathbf{h}(t) \parallel z$, while the $\Gamma_{27}(\Delta M_y, \Delta L_{cx}, \Delta L_{bz})$ mode is excited by field $\mathbf{h}(t) \parallel y$ or by field $\mathbf{h}(t) \parallel z$ for $\mathbf{E}^0 \parallel y$ (see Table 2). In this case, the vibrational variables of the Γ_{16} mode are defined by formulas (48) and (49) after the substitutions $\Delta L_{bz} \longrightarrow \Delta L_{az}$, $E_y^0 \longrightarrow E_x^0$, $q^+ \longrightarrow q^-$, and $\omega_{\rm AQ}(+) \longrightarrow \omega_{\rm AQ}(-)$, while the Γ_{27} mode can be obtained from Γ_{16} under the action of symmetry element $4'_{2z}$ on it.

Finally, the quasi-antiferromagnon (electrically inactive) mode $\Gamma_{34}(\Delta M_z, \Delta L_{ax}, \Delta L_{by})$ is excited by magnetic field $\mathbf{h}(t) \parallel z$ or by the simultaneous action of fields $\mathbf{E}^0 \perp z$ and $\mathbf{h}(t) \perp z$ (see Table 2). The variables of this mode are calculated by formulas (50) and (51) using the substitutions $\Delta L_{ay} \longrightarrow \Delta L_{ax}$, $\Delta L_{bx} \longrightarrow \Delta L_{by}$, $h_x \longleftrightarrow h_y$, $\omega \longrightarrow -\omega$, $K^+ \longrightarrow K^-$, and $\omega_{\rm AF}(+) \longrightarrow \omega_{\rm AF}(-)$, where $\omega_{\rm AF}(-)$ is the natural frequency of the Γ_{34} mode, which is defined by the formula

$$\omega_{\rm AF}(-) = \omega_0\sqrt{\tilde{A}_M K^-}. \quad (60)$$

Magnetization $\mathbf{M}(t)$ (44) associated with oscillations $\Delta\mathbf{M}$ and $\Delta\mathbf{L}_\xi$ ($\xi = a, b$) can also be calculated in this phase by formulas (52)–(59) using the substitutions

$$\begin{gathered}
\chi_{ik}(+) \longrightarrow \chi_{ik}(-), \quad \Delta\chi_{xx}(+) \longrightarrow \Delta\chi_{yy}(-), \\
\Delta\chi_{yy}(+) \longrightarrow \Delta\chi_{xx}(-), \quad \Delta\chi_{zz}(+) \longrightarrow \Delta\chi_{zz}(-), \\
\Delta\chi_{xy}(+) = \Delta\chi_{yx}(+) \longrightarrow \Delta\chi_{xy}(-) = \Delta\chi_{yx}(-), \quad (61) \\
\Delta\chi_{xz}(+) = -\Delta\chi_{zx}(+) \longrightarrow -\Delta\chi_{yz}(-) = \Delta\chi_{zy}(-), \\
\Delta\chi_{yz}(+) = -\Delta\chi_{zy}(+) \longrightarrow -\Delta\chi_{xz}(-) = \Delta\chi_{zx}(-),
\end{gathered}$$

in this case, all plus signs in formulas (53)–(59) are replaced by the minus signs.

The magnetoelectric effects in the $\rm Nd_2CuO_4$ compound were investigated experimentally in [11, 12]. The excitation of the quasi-antiferromagnon mode $\Gamma_{35}(\Delta M_z, \Delta L_{ay}, \Delta L_{bx})$ by a varying electric field $\mathbf{E}(t) \parallel [001]$ was observed in [11] at temperatures $1.5 < T < 50$ K. At a temperature below 10 K, a quadratic dependence ($\Delta\chi_{zz} \propto E_x^2(t)$) of the magnetoelectric part of susceptibility $\Delta\chi_{zz}$ on the amplitude of the quasi-static electric field $\mathbf{E}(t) \parallel [100]$ of frequency $f_E = \omega_E/2\pi = 2.6$ kHz was observed in [12]. The rf magnetic field $\mathbf{h}(t) \parallel [001]$ had a frequency of $f_h = \omega_h/2\pi = 36$ GHz. When a constant electric field $\mathbf{E}^0 \parallel [100]$ was applied in addition, magnetic susceptibility $\Delta\chi_{zz} \propto (E_x^0 + E_x(t))^2$ acquired not only a quadratic, but also a linear dependence on the varying field $\mathbf{E}(t)$($\Delta\chi_{zz} \propto E_x^0 + E_x(t)$).

These experimental results are in good agreement with formulas (29) and (56) if we carry out the substitutions $E_x^0 \longrightarrow E_x^0 + E_x(t)$, $E_y = 0$, and $\omega \longrightarrow \omega_E + \omega_h \approx \omega_h$ in the latter formula.

However, the authors of [12] failed to observe the static linear magnetoelectric effect. This is probably associated with the domain structure of the sample, leading to a decrease in static magnetoelectric susceptibilities $\alpha_{xx}^0 = \alpha_{yy}^0$ and α_{zz}^0, which can be determined from formulas (27) and (29) for $\omega = 0$. According to estimates made by the authors of [12] themselves, these susceptibilities are more than two orders of magnitude smaller than for a polycrystalline Cr_2O_3 sample and do not exceed 5×10^{-7} CGS units. It should be noted that sample poling by an appropriate thermomagnetic treatment might lead to a substantial increase in the values of magnetoelectric susceptibilities. For example, the magnetoelectric susceptibility $\alpha_{zx}^0 \sim 10^{-8}$ CGS units in the antiferromagnetic compound Gd_2CuO_4 increases after magnetoelectric annealing to 3×10^{-5} CGS units at $T = 5$ K [18].

7. CHIRAL PHASE TRANSITIONS

It was mentioned above that the Nd_2CuO_4 compound experiences successive phase transitions $\Gamma_4 \longleftrightarrow \Gamma_5 \longleftrightarrow \Gamma_4$ at temperatures $T_1 = 30$ K and $T_2 = 70$ K [3, 6, 7] between two chiral phases Γ_4 ($Q_z = 1$) and Γ_5 ($Q_z = -1$), which are accompanied by a change in chirality Q_z. A more detailed treatment of these transitions is given below.

To describe these transitions, we write magnetizations $\mathbf{M}_\nu$ of the sublattices in the form

$$M_{\nu x} = M_0 \cos\varphi_\nu, \quad M_{\nu y} = M_0 \sin\varphi_\nu, \tag{62}$$

where φ_ν are azimuth angles of vectors $\mathbf{M}_\nu$ ($\nu = 1, 2, 3, 4$). Taking into account relations (62) in definition (3) of vectors $\mathbf{L}_a$ and $\mathbf{L}_b$, we can represent the anisotropic part of thermodynamic potential Φ_{mag} (9) for $\mathbf{M}^0 = \mathbf{L}_c^0 = 0$ (including the terms with coefficients q_a, q_s, K_1, and K_2 in formula (9)) in the form of the anisotropy energy $\Phi_{\mathrm{an}}(\varphi_\nu)$ as a function of angles φ_ν (we will not write energy $\Phi_{\mathrm{an}}(\varphi_\nu)$ in explicit form since it is quite cumbersome).

Testing energy $\Phi_{\mathrm{an}}(\varphi_\nu)$ for minimum, we find that the stability boundaries of the $\Gamma_4(L_{ax}^0, L_{by}^0)$ phase for values of angles $\varphi_1 = -\varphi_2 = \pi/4$, $\varphi_3 = -\varphi_4 = 3\pi/4$ corresponding to this phase are defined by the inequalities

$$\tilde{q}_a + q_s < 0, \tag{63}$$

$$\frac{1}{2}(K_2 - K_1) - q_s > 0, \tag{64}$$

while the stability of the $\Gamma_5(L_{ay}^0, L_{bx}^0)$ phase for values of angles $\varphi_1 = -\varphi_3 = \pi/4$, $\varphi_2 = -\varphi_4 = 3\pi/4$ corresponding to this phase are defined by the inequalities

$$\tilde{q}_a - q_s > 0, \tag{65}$$

$$\frac{1}{2}(K_1 - K_2) - q_s > 0. \tag{66}$$

In accordance with relation (9), the anisotropy energy Φ_{an} in the stability region of phases $\Gamma_4(L_{ax}^0, L_{by}^0)$ and $\Gamma_5(L_{ay}^0, L_{bx}^0)$ is a negative quantity; consequently, the relativistic constants $q_s < 0$, $K_1 < 0$, and $K_2 < 0$, while the chirality coefficient $\tilde{q}_a$ reverses its sign under phase transitions in accordance with formulas (63) and (65).

Phase transitions $\Gamma_4 \longleftrightarrow \Gamma_5 \longleftrightarrow \Gamma_4$ between chiral phases Γ_4 and Γ_5 occur, according to experiments [6, 7] at two temperatures $T_1^0 = 30$ K and $T_2^0 = 70$ K. At these points of the transition, the quantity

$$a(T) = \tilde{q}_a + \frac{1}{2}(K_1 - K_2) \tag{67}$$

corresponding to the equality of the anisotropy energies of two chiral phases ($\Phi_{\mathrm{an}}(\Gamma_4) = \Phi_{\mathrm{an}}(\Gamma_5)$) vanishes. In the vicinity of transition temperatures T_1^0 and T_2^0, the $a(T)$ dependence (67) can be represented in the form

$$a(T) \approx \lambda_1(T - T_1^0) \quad \text{and} \quad a(T) \approx \lambda_2(T_2^0 - T), \tag{68}$$

respectively, where $\lambda_1 > 0$ and $\lambda_2 > 0$ are the coefficients determining the rate of variation of quantity $a(T)$ at $T = T_1^0$ and $T = T_2^0$.

We can choose a single inequality from (63), (64) or (65), (66), which determines the stability boundary of the Γ_4 or Γ_5 phase. It depends on the sign of the chiral constant $\tilde{q}_a(T_{1,2}^0)$ defined at the phase-transition points. For $\tilde{q}_a(T_{1,2}^0) > 0$, the stability boundaries of the phases are determined by inequalities (63) and (66), respectively, while $\tilde{q}_a(T_{1,2}^0) < 0$, the stability boundaries are determined by inequalities (64) and (65).

Taking into account relations (67) and (68) in inequalities (63)–(66), we find that the Γ_4 and Γ_5 phases are stable at temperatures of

$$T < T_1^+ \quad \text{and} \quad T > T_2^+ \quad (\text{phase } \Gamma_4^+), \tag{69}$$

$$T_1^- < T < T_2^- \quad (\text{phase } \Gamma_5^-), \tag{70}$$

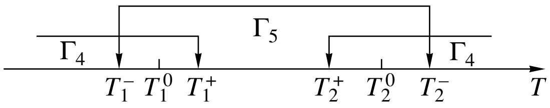


Fig. 3. Temperature hysteresis loops for successive chiral phase transitions $\Gamma_4 \longleftrightarrow \Gamma_5 \longleftrightarrow \Gamma_4$ in Nd_2CuO_4 compound.

where

$$T_1^{\pm} = T_1^0 \pm \Delta_1 \text{ and } T_2^{\pm} = T_2^0 \mp \Delta_2$$

are the stability loss temperatures for chiral phases Γ_4^+ and Γ_5^- (Fig. 3).

The temperature hysteresis loops of the transitions, $\Delta T_1 = T_1^+ - T_1^- = 2\Delta_1$ and $\Delta T_2 = T_2^- - T_2^+ = 2\Delta_2$, are determined by the quantities

$$\Delta_{1,2} = \frac{1}{\lambda_{1,2}}(|q_s| - |q_a|)_{T = T^0_{1,2}}. \quad (71)$$

Thus, successive chiral phase transitions $\Gamma_4 \longleftrightarrow \Gamma_5 \longleftrightarrow \Gamma_4$ occurring at temperatures T_1^0 and T_2^0 are first-order transitions with temperature hysteresis loops $\Delta T_{1,2} = 2\Delta_{1,2}$. Features of these transitions associated with chirality and variation of resonance frequencies at the transition points will be considered in Section 9.

8. EXCITATION OF NMR BY AN ELECTRIC FIELD

The effects of antimagnon and quasi-antiferromagnon excitation by a varying electric field $\mathbf{E}(t)$ considered above are associated with the magnetoelectric and antiferroelectric interactions and usually correspond to the microwave range (or, probably, even higher frequencies).

In the lower (radiofrequency) range, these interactions can be manifested in the effect of resonance excitation of nuclear spins (nuclear magnetoelectric resonance, NMER) by a varying electric field $\mathbf{E}(t)$ [13, 17, 19, 20]. This effect takes place if the frequency of the field $\mathbf{E}(t)$ building up oscillations of vectors $\mathbf{L}$ or $\mathbf{M}$ due to the magnetoelectric interaction is close to the NMR frequency

$$\omega_{n\nu} = \gamma_n H_{n\nu}, \quad (72)$$

determined by the constant component of the hyperfine interaction field in the electron subsystem,

$$H_{n\nu} = FM_\nu^0; \quad (73)$$

γ_n is the nuclear magnetomechanical ratio, F is the hyperfine interaction constant, and M_ν^0 is the magnetization of the νth sublattice in the ground state.

Let us first consider the excitation of NMR by the electric field $\mathbf{E}(t)$ in the $\Gamma_4(L_{ax}^0, L_{by}^0)$ phase with the exchange magnetic structure I characterized by chirality $Q_z = 1$. For this purpose, we supplement thermodynamic potential Φ (11) with the terms

$$\begin{aligned}\Phi_{\mathrm{ME}}^n = {} & -R^n[E_x(\mathbf{L}_a \cdot \mathbf{m}) + E_y(\mathbf{L}_b \cdot \mathbf{m})] \\ & - s_3^n m_z E_z(L_{ax} + L_{by}),\end{aligned} \quad (74)$$

$$\begin{aligned}\Phi_{\mathrm{AFE}}^n = {} & -G^n[E_x(\mathbf{L}_b \cdot \mathbf{l}_c) + E_y(\mathbf{L}_a \cdot \mathbf{l}_c)] \\ & - f_3^n l_{cz} E_z(L_{ay} + L_{bx}),\end{aligned} \quad (75)$$

$$\Phi_{\mathrm{HF}} = -\frac{F}{4}(\mathbf{M} \cdot \mathbf{m} + \mathbf{L}_a \cdot \mathbf{l}_a + \mathbf{L}_b \cdot \mathbf{l}_b + \mathbf{L}_c \cdot \mathbf{l}_c), \quad (76)$$

where Φ_{ME}^n and Φ_{AFE}^n describe the nuclear magnetoelectric and antiferroelectric interactions and Φ_{HF} is the hyperfine interaction of nuclear magnetizations $\mathbf{m}_\nu$ of the sublattices with vectors $\mathbf{M}$ and $\mathbf{L}_\xi$ ($\xi = a, b, c$). Here, $\mathbf{m}$ and $\mathbf{l}_\xi$ are the nuclear vectors of ferromagnetism and antiferromagnetism, which are connected with $\mathbf{m}_\nu$ ($\nu = 1, 2, 3, 4$) through relations analogous to (3).

It should be noted that we have ignored in expressions (74) and (75) some relativistic terms similar to those in formulas (6) and (7) with coefficients s_1, s_2 and f_1, f_2. The inclusion of these terms would only lead to an insignificant renormalization of constants R^n and G^n appearing with the scalar products of vectors $\mathbf{L}$ with $\mathbf{m}$ and $\mathbf{L}$ with $\mathbf{l}$; analogously to the exchange in the electron subsystem, we can assume that these terms exceed anisotropic interactions.

In accordance with relation (73), constant fields $\mathbf{H}_{n\nu}$ of hyperfine interaction in phase $\Gamma_4(L_{ax}^0, L_{by}^0)$ have the components

$$\begin{aligned}H_{n1}^x &= H_{n2}^x = -H_{n3}^x = -H_{n4}^x = (\pm)FM_0/\sqrt{2}, \\ H_{n1}^y &= -H_{n2}^y = H_{n3}^y = -H_{n4}^y = (\pm)FM_0/\sqrt{2}.\end{aligned}$$

These fields correspond to constant nuclear magnetizations $\mathbf{m}_\nu^0 = \chi_n^0 \mathbf{H}_{n\nu}$ of the sublattices with component

$$\begin{aligned}m_{1x}^0 &= m_{2x}^0 = -m_{3x}^0 = -m_{4x}^0 \\ &= (\pm)\chi_n^0 FM_0/\sqrt{2}, \\ m_{1y}^0 &= -m_{2y}^0 = m_{3y}^0 = -m_{4y}^0 \\ &= (\pm)\chi_n^0 FM_0/\sqrt{2},\end{aligned} \quad (77)$$

where χ_n^0 is the static nuclear magnetic susceptibility.

Taking into account expressions (77) in the definition of vectors **m** and $\mathbf{l}_\xi$ (using formulas analogous to (3)), we find that the nuclear magnetization vector $\mathbf{m}^0 = 0$ in the ground state, and the nuclear antiferromagnetism vectors $\mathbf{l}_\xi^0$ have the components

$$l_{ax}^0 = l_{by}^0 = (\pm)\sqrt{8}\chi_n^0 F M_0 \equiv (\pm)\chi_n^0 F L. \tag{78}$$

Let us now separate the quadratic form Φ_2^n from the nuclear part of the thermodynamic potential

$$\Phi^n = \Phi_{\mathrm{ME}}^n + \Phi_{\mathrm{AFE}}^n + \Phi_{\mathrm{HF}}. \tag{79}$$

Taking into account relations (78) as well as the equimodularity conditions for nuclear vectors **m** and $\mathbf{l}_\xi$ (analogous to conditions (14) and (15)), which is valid in the linear theory, we obtain[2]

$$\begin{aligned}\Phi_2^n(\pm) = {} & -(\tilde{G}^n - R^n)(E_x l_{cy} + E_y l_{cx}) - 2\tilde{s}_3^n m_z E_z \\ & - \frac{F}{4}(\Delta M_z m_z + \Delta L_{az} l_{az} \\ & + 2\Delta L_{bx} l_{bx} + 2\Delta L_{by} l_{by} + \Delta L_{bz} l_{bz} \\ & + 2\Delta L_{cx} l_{cx} + 2\Delta L_{cy} l_{cy} + \Delta L_{cz} l_{cz}),\end{aligned} \tag{80}$$

where

$$\tilde{G}^n - \tilde{R}^n = (\pm)(G^n - R^n)L, \quad \tilde{s}_3^n = (\pm)s_3^n L.$$

The NMER signals can be registered from the Q factor by determining the heat loss [17]

$$Q^n = -\overline{\mathbf{P}^n \dot{\mathbf{E}}}. \tag{81}$$

Here, $\mathbf{P}^n$ is the nuclear part of the electric polarization associated with oscillations of vectors **m** and $\mathbf{l}_\xi$. The expression for $\mathbf{P}^n$ can be determined from the definition of effective polarization

$$\mathbf{P} = -\frac{\partial \tilde{\Phi}_2}{\partial \mathbf{E}}, \tag{82}$$

where $\tilde{\Phi}_2 = \Phi_2 + \Phi_2^n$ is the quadratic form of the total thermodynamic potential. Taking into account expressions (24) for Φ_2 and (80) for Φ_2^n in Eq. (82), we obtain

$$P_{x,y} = (\tilde{G} - \tilde{R})\Delta L_{cy,x} + (\tilde{G}^n - \tilde{R}^n) l_{cy,x}, \tag{83}$$

$$P_z = 2(\tilde{s}_3 \Delta M_z + \tilde{s}_3^n m_z). \tag{84}$$

Since we are interested in frequencies

$$\omega \approx \omega_n \ll \omega_{\mathrm{AQ}}, \omega_{\mathrm{AF}}, \tag{85}$$

which are much lower than the natural frequencies ω_{AQ} and ω_{AF} of quasi-antiferromagnon vibrations, we can use the quasi-equilibrium approximation. In this case, variables $\Delta\mathbf{L}_c$ and $\Delta\mathbf{M}$ can be determined directly from the requirement of the energy minimum:

$$\frac{\partial \tilde{\Phi}_2}{\partial \mathbf{L}_c} = 0, \quad \frac{\partial \tilde{\Phi}_2}{\partial \mathbf{M}} = 0.$$

This gives

$$\Delta L_{cx,y} = \frac{\tilde{G} - \tilde{R}}{\tilde{A}} E_{y,x} + \frac{F}{2\tilde{A}} l_{cx,y}, \tag{86}$$

$$\Delta M_z = \frac{2\tilde{s}_3}{\tilde{A}_M} E_z + \frac{F}{4\tilde{A}_M} m_z. \tag{87}$$

Finally, to determine the resonance response to field $\mathbf{E}(t)$ in the form of effective polarization (83), (84), we must solve the equations of motion for **m** and $\mathbf{l}_c$ (in the linear and nonlinear approximation depending on the experimental conditions, i.e., continuous or pulsed mode). It was noted above that we consider the results for the linear case only. For $l_{c\alpha}$ ($\alpha = x, y$) and m_z, we obtain

$$l_{cx,y} = 2\chi_n(\omega)\left[(\tilde{G}^n - \tilde{R}^n) + \frac{\tilde{G} - \tilde{R}}{2\tilde{A}} F\right] E_{y,x}, \tag{88}$$

$$m_z = 8\chi_n(\omega)\left(\tilde{s}_3^n + \frac{\tilde{s}_3}{4\tilde{A}_M} F\right) E_z. \tag{89}$$

Here,

$$\chi_n(\omega) = \chi_n^0 \frac{\omega_n^2}{\omega_n^2 - \omega^2} \tag{90}$$

is the NMR susceptibility and $\omega_n = \gamma_n F M_0$ is the NMR frequency.

Formulas (88) and (89) describe two resonance excitation channels for NMER signals. The terms containing $\tilde{G}^n - \tilde{R}^n$ and $\tilde{s}_3^n$ describe direct excitation of polarization $\mathbf{P}^n(t)$ by oscillations of quantities $l_{c\alpha}$ ($\alpha = x, y$) and m_z (see formulas (83) and (84)). The second terms containing $\tilde{G} - \tilde{R}$ and $\tilde{s}_3$ are responsible for indirect excitation via the hyperfine interaction of vectors $\mathbf{L}_c$ with $\mathbf{l}_c$ and **M** with **m**. Each of these channels is characterized by its own magneto- and antiferroelectric constants.

---

[2] These conditions remain in force for the paramagnetic spin system (including the nuclear subsystem) also if we remain in the framework of the linear theory in nuclear spin oscillations, which will be considered below.

Substituting expressions (86)–(89) into Eqs. (83) and (84), we find that the nuclear part of the polarization is defined as

$$P_\alpha^n = 2\chi_n(\omega)\left[\tilde{G}^n - \tilde{R}^n + \frac{\tilde{G}-\tilde{R}}{2\tilde{A}}F\right]^2 E_\alpha, \qquad \alpha = x, y, \tag{91}$$

$$P_z^n = 8\chi_n(\omega)\left(\tilde{s}_3^n + \frac{\tilde{s}_3}{4\tilde{A}_M}F\right)^2 E_z. \tag{92}$$

It is the $\mathbf{P}^n$ polarization with components P_α^n (91) and P_z^n (92), which is associated with the nuclear subsystem and induced by field $\mathbf{E}(t)$, that is a linear NMER signal manifested in the antinode of this field.

Calculating the absorbed energy Q^n (81) taking into account relations (91) and (92), we obtain

$$Q^n = \omega\chi_n''(\omega)\left\{\left[\tilde{G}^n - \tilde{R}^n + \frac{\tilde{G}-\tilde{R}}{2\tilde{A}}F\right]^2 \times(|E_x|^2 + |E_y|^2) + 4\left(\tilde{s}_3^n + \frac{\tilde{s}_3}{4\tilde{A}_M}F\right)^2|E_Z|^2\right\}, \tag{93}$$

where $\chi_n''(\omega)$ is the imaginary part of quantity $\chi_n(\omega)$ (90), which is given by

$$\chi_n''(\omega) = \chi_n^0\frac{2\omega\Gamma_n\omega_n^2}{(\omega_n^2-\omega^2)^2 + (2\omega\Gamma_n)^2}. \tag{94}$$

Here, dissipation is taken into account again through the substitution $\omega \longrightarrow \omega + i\Gamma_n$ (in the Bloch sense).

In the $\Gamma_5(L_{ay}^0, L_{bx}^0)$ phase with the exchange magnetic structure II and chirality $Q_z = -1$, the NMER effect can also be described by formulas of the form (91)–(94) after the following substitution of constants:

$$\tilde{s}_3^n \longrightarrow \tilde{f}_3^n = (\pm)f_3^n L,$$

$$\tilde{s}_3 \longrightarrow \tilde{f}_3 = (\pm)f_3 L, \quad \tilde{A}_M \longrightarrow \tilde{A}_c.$$

Consequently, as should be expected for a first-order phase transition, the NMER signal changes its value jumpwise during a chiral transition (true, the jump appears in formula (91) only if we additionally take into account the relativistic magneto- and antiferroelectric interactions).

9. DISCUSSION

The main concept adopted in this study on the basis of [7, 11, 12], in which the existence of the magnetoelectric and antiferroelectric interactions in Nd_2CuO_4 was established, is the assumption that crystal-chemistry group $P4_2/mnm \equiv D_{4h}^{14}$ corresponds to the symmetry of this cuprate (it differs from the body-centered group $I4/mmm \equiv D_{4h}^{17}$, which was purportedly ascertained by some other authors for neodymium cuprate [1–4]). This allowed us to study in detail the dynamics of the Nd_2CuO_4 compound using the concept of a chiral doublet and a phase transition between the components of this doublet.

Another important aspect is the assumption concerning the possibility of existence of a domain structure, in which domains equivalent in energy and other parameters differ only in the signs of the antiferromagnetism vectors and in the signs of the magnetoelectric magnetizations associated with them. This may nullify (or at least reduce) the total magnetoelectric magnetization (magnetoelectric effect linear in field **E**) and the field-induced magnetoelectric susceptibility associated with it for nonzero absorption of the energy of the electric field, which is a quadratic function of the field.

Finally, an appropriate representation of the magnetoelectric interaction in form (43) has made it possible to explain the effect quadratic in **E** observed in [12], which is determined by the linear magnetoelectric interaction and is independent of the above-mentioned domain structure when field **E** is applied in addition to a varying field $\mathbf{H} \equiv \mathbf{h}(t)$.

Thus, in the four-sublattice antiferromagnet Nd_2CuO_4 with the crystal structure described by group (2), the magnetic Cu^{2+} ions occupy a noncentrosymmetric position, which explains the existence of both magnetoelectric (6) and the antiferroelectric (7) interaction. The main terms are of the exchange type. In addition, noncollinear magnetic structures (phases) I and II of the square (or rectangular cross) type in neodymium cuprate form exchange doublets with different chiralities and magnetic symmetries. For example, the symmetry of the $\Gamma_4(L_{ax}^0, L_{by}^0)$ phase with exchange structure I and chirality $Q_z = 1$ is determined by the magnetic group $\bar{1}'4_{2z}2_{1x}$, while the symmetry of the $\Gamma_5(L_{ay}^0, L_{bx}^0)$ phase with exchange structure II and chirality $Q_z = -1$ is determined by the magnetic group $\bar{1}'4'_{2z}2'_{1x}$.

One more important remark is appropriate here. In this study, we assume that noncollinearity of the structures (I and II) is associated with the exchange (see inequalities (13)). At the same time, for the other version (1) of the crystal-chemistry group, it was proved by some authors that the noncollinearity of the structure might be due to the pseudodipole interaction [3, 4]. At any rate, the role of this interaction should be analyzed in greater detail in our model also, although we agreed not to discuss the specific microscopic mechanisms of interactions.

Each of the phases considered here, e.g., in phase Γ_4 (or Γ_5), is characterized by four modes of spin oscillations. One of these modes, Γ_{48} (20) (or Γ_{58} (31)), is an antimagnon mode and has a frequency $\omega_L(+)$ (47) (or $\omega_L(-)$ (37)), which depends on the difference (or sum) of the biquadratic exchange constants D and chirality constant $\tilde{q}_a$ (the latter is of semiexchange–semirelativistic origin). (It should be recalled that the plus and minus signs in the arguments or superscripts of physical quantities indicate that they correspond to the Γ_4 or Γ_5 phases with chiralities $Q_z = +1$ and $Q_z = -1$, respectively.) The other two quasi-antiferromagnon modes, Γ_{17} (21) and Γ_{26} (22) (or Γ_{16} and Γ_{27} (32)) are degenerate and have frequency $\omega_{AQ}(+)$ (28) (or $\omega_{AQ}(-)$ (41)), which is determined by the geometric mean value of exchange constant $\tilde{A}$ and renormalized chirality coefficient q^+ (or q^-). Finally, the fourth mode Γ_{35} (23) (or Γ_{34} (33)) has a frequency $\omega_{AF}(+)$ (30) (or $\omega_{AF}(-)$ (60)) characterized by the geometrical mean value of the exchange constant $\tilde{A}_M$ and the linear combination K^+ (or K^-) consisting of relativistic magnetic anisotropy constants.

It is shown in Table 2 that if $\mathbf{E}^0 = 0$, some of these modes cannot be excited by electric field $\mathbf{E}(t)$ or magnetic field $\mathbf{h}(t)$ (e.g., the antimagnon mode $\Gamma_{48}(\Delta L_{ax}, \Delta L_{by}, \Delta L_{cz})$ cannot be excited by either of these fields). However, in addition to the cases considered in [7], we also considered the case when $\mathbf{E}^0 \neq 0$, for which resonance is possible (for certain directions of the fields) for all modes under investigation (see the fourth column in Table 2). The natural frequencies remain approximately unchanged (as in the case when $\mathbf{E}^0 = 0$).

It has been established experimentally that Nd_2CuO_4 exhibits successive phase transitions $\Gamma_4 \longleftrightarrow \Gamma_5 \longleftrightarrow \Gamma_4$, associated with a change in chirality at temperatures $T_1 = 30$ K and $T_2 = 70$ K [6, 7]. It was found that these are first-order phase transitions of the order–order type.[3]

Let us now analyze the behavior of resonance frequencies during the phase transition under study by using Table 2. A simple analysis shows that two situations are possible in this case.

1. If chirality coefficient $\tilde{q}_a(T^0_{1,2}) > 0$ at points $T = T^0_{1,2}$ of the phase transition (see Fig. 3), the quantities

$$\tilde{q}_a + q_s\big|_{T = T^+_{1,2}} = 0 \quad \text{and} \quad K^-\big|_{T = T^-_{1,2}} = 0 \qquad (95)$$

vanish at critical temperatures $T = T^+_{1,2}$ and $T = T^-_{1,2}$, which are equal to the temperatures of stability loss of phases $\Gamma_4(+)$ and $\Gamma_5(-)$. The most interesting fact here is that frequency $\omega_{AF}(-)$ (60) of the quasi-antiferromagnon mode Γ_{34} (33) vanishes at points $T = T^-_{1,2}$ (analogously to the case of a soft mode in a second-order phase transition),

$$\omega^-_{AF}(T^-_{1,2}) = 0. \qquad (96)$$

Naturally the Γ_5 phase vanishes in this case, giving way to the Γ_4 phase with a nonzero frequency at these points:

$$\omega^+_{AF}(T^-_{1,2}) = 2\omega_0\sqrt{2\tilde{A}_M|q_s|}. \qquad (97)$$

For all other frequencies at the stability-loss points $T = T^-_{1,2}$ (second condition in (95)) as well as at points $T = T^+_{1,2}$ (corresponding to the first condition in (95)), changes (discontinuity) typical of first-order transitions take place with simultaneous change in the phase (however, the frequency of one of the phases does not vanish in this case). The corresponding results can be obtained using Table 2.

2. If the chirality coefficient $\tilde{q}_a(T^0_{1,2}) < 0$, temperatures $T = T^+_{1,2}$ and $T = T^-_{1,2}$ are determined by the equalities

$$K^+\big|_{T = T^+_{1,2}} = 0 \quad \text{and} \quad \tilde{q}_a - q_s\big|_{T = T^-_{1,2}} = 0, \qquad (98)$$

respectively. In this case, the natural frequency $\omega_{AF}(+)$ (30) of the quasi-antiferromagnon mode Γ_{35} (23) vanishes at temperatures $T = T^+_{1,2}$:

$$\omega^+_{AF}(T^+_{1,2}) = 0. \qquad (99)$$

The corresponding frequency $\omega^-_{AF}(T^+_{1,2})$ in the emerging phase Γ_5 (instead of Γ_4) assumes a finite value (97) again. The statements concerning other modes and analogous to those formulated in the previous case remain in force.

---

[3] In [21], magnetic structures with frustration are considered (i.e., it is impossible to obtain collinear antiferromagnetism in a system of antiferromagnetically interacting spins). Frustration leads to the emergence of exchange-noncollinear (chiral) structures, e.g., of the triangle type in a system of three spins in the same position of multiple points. In our case of four one-position spins, frustration is absent since the spins can be ordered in an antiferromagnetically collinear manner. Exchange-noncollinear chiral magnetic structures (see Fig. 2) emerge from the requirement of energy minimum. We mention Kawamura's review [21] here only due to the fact that the possibility of existence of a new class of phase transitions associated with chirality is discussed in this review. However, the transitions from a paramagnetic structure to an ordered antiferromagnetic chiral structure was studied in [21], while we consider a phase transition from one ordered state to another with a change in chirality in the framework of the conventional Landau theory. Consequently, the special universal phase transition considered in [21] and accompanying the emergence of a new property (chirality) does not take place in our case.

The following circumstance is worth noting. It is expedient to experimentally measure the magnetoelectric contributions $\Delta\chi_{ik}$ (54)–(59), (61) to the magnetic susceptibility at low frequencies $\omega \ll \omega_0$ in the vicinity of critical temperatures $T^+_{1,2}$ or $T^-_{1,2}$, where the natural frequency $\omega_{AF}(+)$ or $\omega_{AF}(-)$ of the quasi-antiferromagnon mode Γ_{35} (23) or Γ_{34} (33) may vanish at a temperature equal to one of these temperatures. At these temperatures, the component $\Delta\chi_{ik}$ associated with oscillations of these soft mode have the highest value. In particular, for electric field $\mathbf{E}^0 \parallel x$, from expression (54)–(59) and (61), we obtain

$$\Delta\chi^+_{yy}(T^+_{1,2}) = \Delta\chi^-_{xx}(T^-_{1,2}) = -\frac{\omega_0^2}{\omega^2}R'^2\tilde{A}_M(E^0_x)^2, \quad (100)$$

$$\begin{gathered}\Delta\chi^+_{yz}(T^+_{1,2}) = -\Delta\chi^+_{zy}(T^+_{1,2}) \\ = -\Delta\chi^-_{xz}(T^-_{1,2}) = \Delta\chi^-_{zx}(T^-_{1,2}) = i\frac{\omega_0}{\omega}R'E^0_x.\end{gathered} \quad (101)$$

In addition, for $T = T^+_{1,2}$ (when $\omega_{AF}(+) = 0$ or $\omega_{AF}(-) = 0$), magnetic susceptibility $\chi^\pm_{zz} = 0$ (53), which allows us to eliminate this Zeeman contribution in the measurement of magnetoelectric correction $\Delta\chi^\pm_{zz}$ (56) in the vicinity of chiral phase transitions.

Thus, experimental investigation of the temperature dependences of the natural frequencies of spin oscillations would make it possible to determine the type and stability region of chiral phases in successive phase transitions $\Gamma_4 \longleftrightarrow \Gamma_5 \longleftrightarrow \Gamma_4$ in Nd_2CuO_4. It should be borne in mind, however, that antimagnon mode Γ_{48} (20) of the Γ_4 phase and quasi-antiferromagnon mode Γ_{34} (33) of the Γ_5 phase are not electrically active (i.e., are not excited by a varying electric field $\mathbf{E}(t)$). On the other hand, these and all the remaining modes can be excited by a varying magnetic field $\mathbf{h}(t)$ in the presence of a constant electric field $\mathbf{E}_0$ = const (see Table 2). It should be noted that electrically active modes Γ_{58} (31) and Γ_{35} (23) are excited by electric field $\mathbf{E}(t) \parallel z$, while modes Γ_{17} (21), Γ_{26} (22), and Γ_{16}, Γ_{27} (32) are excited by field $\mathbf{E}(t) \perp z$.

NMER experiments on the Cu^{63} and Cu^{65} nuclei can also be used for studying the magneto- and antiferroelectric interaction. The highest level of NMER signals in neodymium cuprate should be expected when these interactions are excited by an electric field $\mathbf{E}(t) \perp z$. In this case, polarization P^n_α ($\alpha = x, y$) (91) and the amount of heat Q^n (93) are determined by the exchange part of the magnetoelectric and antiferroelectric interactions. The electron and nuclear parts of these interactions are characterized by the difference of constants $\tilde{G} - \tilde{R}$ and $\tilde{G}^n - \tilde{R}^n$. The values of these constants have not yet been determined.

As an added note, the ratio of constants $(\tilde{G} - \tilde{R})/\tilde{A}$ appearing in the expression for P^n_α (91) and Q^n (93) could be determined from experimental measurements of magnetoelectric susceptibility $\hat{\alpha}$ by studying the dynamic magnetoelectric effect emerging, for example, upon excitation of quasi-antiferromagnon modes Γ_{17} (21) and Γ_{26} (22) of the Γ_4 phase by electric field $\mathbf{E}(t) \perp z$. In this case, magnetoelectric susceptibilities α_{xx} (27) and α_{yy} at low frequencies $\omega^2 \ll \omega^2_{AQ}(+)$ are determined precisely by this ratio:

$$\alpha_{xx} = \alpha_{yy} = -\frac{\tilde{G} - \tilde{R}}{\tilde{A}}. \quad (102)$$

Thus, the above arguments confirm once again the urgent need for experiments on antiferroelectric and nuclear magnetoelectric resonances in the chiral antiferromagnet Nd_2CuO_4.

ACKNOWLEDGMENTS

The authors thank N.G. Bebenin, M.I. Kurkin, and Yu.N. Skryabin for fruitful discussions.

This study was financed by the Russian Foundation for Basic Research (project no. 02-02-16440).

REFERENCES

1. D. Petitgrand, A. H. Moudden, P. Galez, *et al.*, J. Less-Common Met. **164–165**, 768 (1990).
2. O. Kondo, M. Ono, T. Yosida, *et al.*, J. Magn. Magn. Mater. **90–91**, 79 (1990).
3. G. R. Sachidanandam, T. Yildirim, A. B. Harris, *et al.*, Phys. Rev. B **56**, 260 (1997).
4. D. Petitgrand, S. V. Maleyev, Ph. Bourges, *et al.*, Phys. Rev. B **59**, 1079 (1999).
5. D. A. Yablonskiĭ and V. N. Krivoruchko, in *Problems in Physical Kinetics and Physics of Solid State* (Naukova Dumka, Kiev, 1990) [in Russian].
6. S. Skanthakumar, H. Zhang, T. W. Clinton, *et al.*, Physica C (Amsterdam) **160**, 124 (1989).
7. V. A. Blinkin, I. M. Vitebskiĭ, O. D. Kolotiĭ, *et al.*, Zh. Éksp. Teor. Fiz. **98**, 2098 (1990) [Sov. Phys. JETP **71**, 1179 (1990)].
8. E. A. Turov, Zh. Éksp. Teor. Fiz. **115**, 1386 (1999) [JETP **88**, 766 (1999)].
9. E. A. Turov, A. V. Kolchanov, V. V. Men’shenin, *et al.*, *Symmetry and Physical Properties of Antiferromagnetics* (Fizmatlit, Moscow, 2001) [in Russian].

10. E. A. Turov, Usp. Fiz. Nauk **164**, 325 (1994) [Phys. Usp. **37**, 303 (1994)].
11. A. I. Smirnov, S. N. Barilo, and D. I. Zhigunov, Zh. Éksp. Teor. Fiz. **100**, 1690 (1991) [Sov. Phys. JETP **73**, 934 (1991)].
12. A. I. Smirnov and I. N. Khlyustikov, Zh. Éksp. Teor. Fiz. **105**, 1040 (1994) [JETP **78**, 558 (1994)].
13. I. F. Mirsaev and E. A. Turov, Zh. Éksp. Teor. Fiz. **124**, 338 (2003) [JETP **97**, 305 (2003)].
14. E. A. Turov, A. V. Kolchanov, V. V. Men'shenin, *et al.*, Usp. Fiz. Nauk **168**, 1303 (1998) [Phys. Usp. **41**, 1191 (1998)].
15. Yu. A. Izyumov and N. A. Chernoplekov, in *Neutron Spectroscopy*, Ed. by R. P. Ozerov (Énergoatomizdat, Moscow, 1983; Consultants Bureau, New York, 1994), Vol. 3.
16. V. G. Bar'yakhtar, I. M. Vitebskiĭ, and D. A. Yablonskiĭ, Zh. Éksp. Teor. Fiz. **76**, 1381 (1979) [Sov. Phys. JETP **49**, 703 (1979)].
17. E. A. Turov and V. V. Nikolaev, Usp. Fiz. Nauk **175** (5) (2005).
18. A. I. Smirnov and I. N. Khlyustikov, Zh. Éksp. Teor. Fiz. **108**, 706 (1995) [JETP **81**, 384 (1995)].
19. M. I. Kurkin, V. V. Leskovets, V. V. Nikolaev, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **45**, 653 (2003) [Phys. Solid State **45**, 685 (2003)].
20. M. I. Kurkin, I. F. Mirsaev, and E. A. Turov, Zh. Éksp. Teor. Fiz. **125**, 1144 (2004) [JETP **98**, 1002 (2004)].
21. H. Kawamura, J. Phys.: Condens. Matter **10**, 4707 (1998).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 100, No. 6, 2005, pp. 1112–1120.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 6, 2005, pp. 1262–1271.
Original Russian Text Copyright © 2005 by Meĭlikhov, Farzetdinova.

Exchange Bias of Hysteresis Loop in the Ising Two-Dimensional Ferromagnet–Antiferromagnet Structure

E. Z. Meĭlikhov and R. M. Farzetdinova

Russian Research Center Kurchatov Institute, Moscow, 123182 Russia

e-mail: meilikhov@imp.kiae.ru

Received October 13, 2004

Abstract—A simple analytical model is developed to explain the phenomenon of exchange bias of the hysteresis loop in a two-dimensional ferromagnet–antiferromagnet bilayer. A solution of the magnetic relaxation equation is obtained within the framework of the generalized mean field theory, which describes the shape of the hysteresis loop and shows its dependence on the properties of a model interface in the system under consideration. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The experimental magnetic hysteresis loop of a ferromagnetic substance (plotted in the coordinates of external magnetic field H_e versus magnetic moment) is usually symmetric. However, the hysteresis loop in a system comprising a ferromagnet (FM) in contact with an antiferromagnet (AFM) may exhibit a shift along the field axis [1, 2]. As a result, the center of symmetry of the loop is displaced to a point with $H_e = H_{\rm EB} \neq 0$. In typical systems of this kind (Co–CoO, Fe–FeF$_2$), such a displacement is quite large and can even be comparable to the coercive force. This phenomenon, called exchange bias (EB), takes place in systems where the Curie temperature ($T_{\rm C}$) of the FM exceeds the Neél temperature ($T_{\rm N}$) of the AFM. The EB phenomenon is observed upon cooling of the system in an external magnetic field from an initial temperature in the interval $T_{\rm N} < T < T_{\rm C}$ to a final temperature $T < T_{\rm N}$. If the direction of the field in which the system is cooled is considered positive, the hysteresis loop shifts in the negative direction: $H_{\rm EB} < 0$. A considerable growth in the interest in this phenomenon stems from the possibility of its various practical applications [1, 3].

It is a common opinion that the nature of exchange bias is related to features of the spin structure of an AFM in the vicinity of the interface, namely, to the domain structure stabilized by nonmagnetic defects (impurities, structural imperfections, etc.) [1, 2, 4]. These defects account for the formation of domains in the AFM, while their random distribution results in that the numbers of spins in the two sublattices are not equal for any finite volume in the AFM. As a result, any such volume possesses a finite magnetic moment whose rotation in an external magnetic field may become energetically favorable [4]. In order to provide for minimization of the domain structure energy, the domain walls must pass predominantly through nonmagnetic defects and, as a result, acquire a complicated shape. Although the resulting structure is metastable, the domain walls exhibit pinning that leads to "freezing" of this structure, whereby it does not change (provided that the temperature is sufficiently low) in response to alteration of the magnetic field sign and magnitude (at least in the range of fields typically encountered in the measurements of hysteresis loops).

The first models (for the most part, phenomenological) of exchange bias did not take into account the interaction of spins in the FM, even within the framework of the mean field theory [2]. Subsequent theories treated this phenomenon from the standpoint of the Ising or Heisenberg models. However, these models, while taking this interaction into account, are reduced to numerical calculations of the magnetic state of the system (e.g., by the Monte Carlo method) [5]. The aim of this study is to develop a simple analytical model representing a generalization of the mean field theory for the Ising system of spins in a two-dimensional (2D) FM under the conditions when the interaction of some of these spins is enhanced due to the exchange interaction with spins of the AFM. Such a model is unavoidably rather simplified, but it nevertheless reflects all the main features of the phenomenon under consideration and admits (owing to this very simplicity) an analytical solution, which allows the influence of various factors on the shape of the hysteresis loop to be readily established.

One possible model mechanism of the specific enhancement of the exchange interaction at the FM–AFM interface was described by Illa *et al.* [5]. This model is illustrated in Fig. 1, which shows the AFM layer with a domain and a domain wall (DW), the FM–AFM interface, and spins of the FM (s_i, s_j) and the AFM

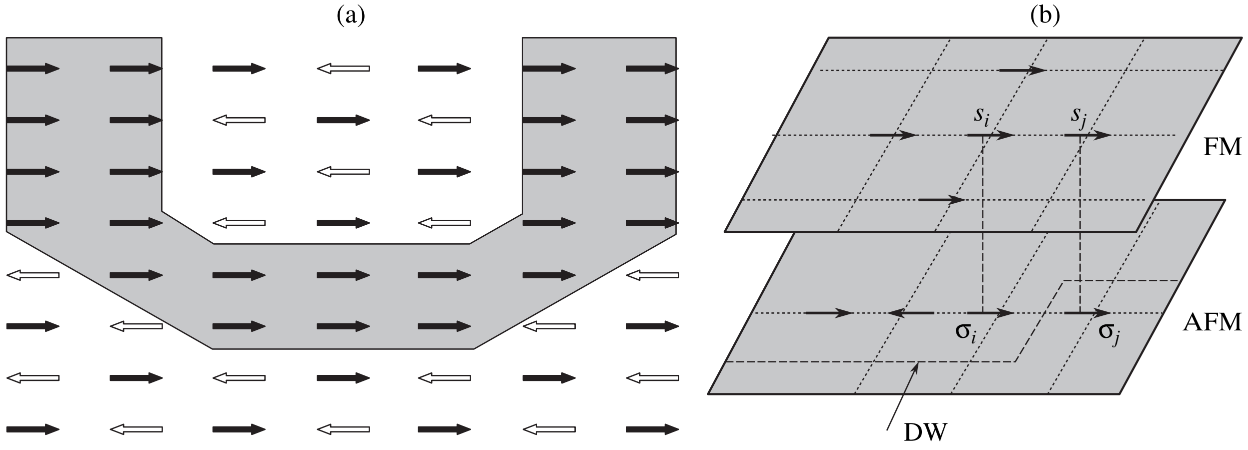


Fig. 1. Schematic diagram showing (a) an AFM layer with a domain wall (DW) and (b) an FM–AFM interface, a domain wall, and coupled spins of the FM (s_i, s_j) and the AFM (σ_i, σ_j) in the vicinity of this interface.

(σ_i, σ_j) in the vicinity of this interface. The FM spins s_i and s_j interact with each other (both directly and indirectly—via the neighboring atoms of the AFM), as well as with the AFM spins. The total interaction energy can be expressed as

$$E_{ij} = E_1 + E_2, \quad E_1 = -(J_2^{F/F} s_i s_j + J_4^{F/AF} s_i s_j \sigma_i \sigma_j),$$
$$E_2 = -J_2^{F/AF} s_i \sigma_i,$$

where $J_2^{F/F}$, $J_2^{F/AF}$, and $J_4^{F/AF}$ are the energy parameters of the bilinear and the four-spin exchange. The energy E_1 can be expressed in terms of an effective constant of the pair interaction between spins s_i and s_j. This constant depends on whether the spin pair occurs far from the AFM domain wall or near this interface. For the spin pair remote from the domain wall, the effective energy is

$$E_1' = -J s_i s_j, \quad J = J_2^{F/F} - J_4^{F/AF} |\sigma_i \sigma_j|,$$

while for the spins near the wall, this energy is

$$E_1'' = -J_E s_i s_j, \quad J_E = J_2^{F/F} + J_4^{F/AF} |\sigma_i \sigma_j|.$$

As can be seen,

$$\frac{J_E}{J} = \frac{J_2^{F/F} + J_4^{F/AF}|\sigma_i \sigma_j|}{J_2^{F/F} - J_4^{F/AF}|\sigma_i \sigma_j|} > 1,$$

which implies that the domain wall in the AFM leads to enhancement of the spin coupling in the FM.

Since the domain contains a large number of lattice sites, the fraction f of the FM spins with such enhanced coupling is small ($f \ll 1$).

It should be emphasized that the above mechanism of enhancement of the spin coupling is just a model. The true mechanism can be quite different: it has only to provide for the appearance of a small fraction of spins with enhanced coupling. This is the only condition necessary for the further analysis.

Subsequent consideration consists of two steps. The first step is related to the random distribution of energies $J_{ij}^{(1)}$ of the pair interaction between the near-wall FM spins described by the function

$$F_J(J_{ij}^{(1)}) = (1-f)\delta(J_{ij}^{(1)} - J) + f\delta(J_{ij}^{(1)} - J_E),$$

as well as to the random distribution of energies $J_{ij}^{(2)}$ of the pair interaction of FM and AFM spins described by the function[1]

$$F_J(J_{ij}^{(2)}) = \frac{1}{2}[\delta(J_{ij}^{(2)} - I) + \delta(J_{ij}^{(2)} + I)],$$

where $I \equiv J_2^{F/AF}$.

Application of the traditional mean field theory leads to significant errors. A more correct result can be obtained by finding and using a distribution function $F(H)$ of the effective local magnetic fields H.

The second step is generalization of the equation describing the relaxation of magnetization of the FM with random magnetic bonds. These interactions are related to the local fields described by the distribution function $F(H)$. The generalized relaxation equation is used for an analysis of the shape and parameters of the hysteresis loop.

It should be noted that the approach developed below is quite general and can be applied not only to the above model of modification of the spin coupling at the

---

[1] A small "frozen" magnetization of an antiferromagnet is mainly due to its spins at antiphase domain walls (see Fig. 1). Consequently, the number of s_i, σ_i pairs with parallel and antiparallel spin orientations can be assumed to be the same; for the same reason, the sign of the interaction constant $J_2^{F/AF}$ is immaterial.

Table 1. Configurations of the magnetic moments of nearest neighbors and their probabilities on a square lattice

Configuration	Effective magnetic field H/h_J	Number of configurations	Probability of configuration (for magnetization j)
↑↑↑↑	+4	$C_4^0 = 1$	$\left(\frac{1+j}{2}\right)^4$
↑↑↑↓	+2	$C_4^1 = 4$	$4\left(\frac{1+j}{2}\right)^3\left(\frac{1-j}{2}\right)$
↑↑↓↓	0	$C_4^2 = 6$	$6\left(\frac{1+j}{2}\right)^2\left(\frac{1-j}{2}\right)^2$
↑↓↓↓	–2	$C_4^3 = 4$	$4\left(\frac{1+j}{2}\right)\left(\frac{1-j}{2}\right)^3$
↓↓↓↓	–4	$C_4^4 = 1$	$\left(\frac{1-j}{2}\right)^4$

FM–AFM interface, but to any other mechanism of such a modification changing the energy of interaction for a small fraction f of spin–spin bonds.

2. GENERALIZED MEAN FIELD MODEL FOR A SQUARE ISING LATTICE

As is well known, the mean field theory does not take into account the correlation of magnetic moments. Another drawback of this theory is the notion of equivalence of all lattice sites, according to which the mean field is assumed to be the same at all sites. Actually, the field varies from one lattice cite to another in a random manner. It would be natural to take this random variation into account and to check how much this generalized mean field theory will increase the accuracy of description. Such an approach was originally developed by Klein and Brout [6] for a system of randomly arranged magnetic dipoles. Later, this approach was thoroughly analyzed by Thomsen *et al.* [7] proceeding from a selected (although without sufficient ground) distribution of the pair interaction energy.

The essence of the generalized model consists in replacing the standard mean field equation

$$j = \tanh\left(\frac{\lambda j}{kT}\right), \tag{1}$$

where j is the reduced magnetization and λ is the mean field constant, by the generalized equation

$$j = \int_{-\infty}^{\infty} \tanh\left(\frac{\mu H}{kT}\right) F(j;\, H) dH. \tag{2}$$

Here, $F(j; H)$ is a function describing the distribution of local magnetic fields H created by all magnetic moments μ except one at the site of location of this magnetic moment in a system (not necessarily regular) with the magnetization j. It should be noted that the standard mean field equation (1) is also equivalent to the equation

$$j = \tanh\left[\frac{\mu \bar{H}(j)}{kT}\right], \quad \bar{H}(j) = \int_{-\infty}^{\infty} HF(j;\, H)dH, \tag{3}$$

which is obtained from the generalized equation (2) by substituting $F(j; H) = \delta[H - \bar{H}(j)]$. According to this approximation, all spins occur in the same field equal to the average field $\bar{H}(j)$.

In order to assess how this generalization improves the theory as compared to the traditional mean field theory, let us calculate the critical temperature and compare the result to the known exact value. Consider a square lattice and let magnetic moments with the two possible orientations, ↑ and ↓, occur at the lattice sites with the probabilities $(1 + j)/2$ and $(1 - j)/2$. If all magnetic bonds are identical ($J_{ij} = J$, or $f = 0$), the energy W of interaction between a certain magnetic moment and its nearest surrounding is the algebraic sum of components with the same absolute value J and the sign determined by the mutual orientation of the “central” moment and its neighbor. Thus, the energy W depends on the configuration of the Ising magnetic moments in the first coordination “sphere.” For a square lattice, there are $2^4 = 16$ such combinations. Their distribution with respect to the effective magnetic field $H = -W/\mu$, as well as the probabilities of various configurations in the system with magnetization j, is given in Table 1 (where $h_J = J/\mu$).

The corresponding distribution function is

$$\begin{aligned} F_0(j;\, H) = \frac{1}{16}[&(1+j)^4\delta(4h_J) \\ &+ 4(1+j)^3(1-j)\delta(2h_J) + 6(1+j)^2(1-j)^2\delta(0) \\ &+ 4(1+j)(1-j)^3\delta(-2h_J) + (1-j)^4\delta(-4h_J)], \end{aligned} \tag{4}$$

where $\delta(h)$ is the delta function taking a nonzero value at $H = h$. Substituting function (4) into Eq. (2), we obtain

$$j^2 = \frac{2\tanh(2K) + \tanh(4K) - 2}{2\tanh(2K) - \tanh(4K)}, \quad K = \frac{\mu h_J}{kT}. \tag{5}$$

This relation describes the temperature dependence of the magnetization in the system under consideration and determines the Curie temperature T_C. This temperature is found from the conditions $j = 0$, which leads to the equation

$$2\tanh(2K) + \tanh(4K) = 2,$$

and yields $K = K_C \approx 0.323$, or

$$kT_C \approx 3.10J. \tag{6}$$

This value is much closer to the exact result ($kT_C \approx 2.27J$) than the value predicted by the mean field theory ($kT_C = 4J$). It should be noted that the field determined by the distribution function (4) is $\overline{H} = 4jh_J$, which leads to the usual mean-field expression for the Curie temperature.

Figure 2 presents the temperature dependences of magnetization, $j(T)$, calculated according to the traditional mean field theory (Eq. (1)), the proposed generalized theory (Eq. (2)), and the exact solution

$$j = [1 - \sinh^{-4}(2K)]^{1/8} \tag{7}$$

of the Ising problem on a square lattice [8]. As can be seen, the result provided by the proposed generalization is much closer than the traditional approximation to the exact solution.

For an FM with zero magnetization ($j = 0$), the distribution function (4) is even with respect to the magnetic field and, hence, the hysteresis loop of the corresponding system is symmetric ($H_{EB} = 0$). Let us generalize the proposed procedure to the case when a 2D ferromagnetic layer comes in contact with an antiferromagnetic layer, so that the fraction of the possible nearest-neighbor configurations with enhanced coupling becomes nonzero ($0 < f \ll 1$). In this case, the effective magnetic moments of all configurations presented in Table 1 will equiprobably change by the value $\pm I/J$ as a result of the additional interaction with the AFM spins far from the domain wall. This interaction modifies the distribution function (4), whereby the arguments of all delta functions change by $\pm I/J$. However, the most important (from the standpoint of a possible shift of the hysteresis loop) property of the distribution function is retained: it is still symmetric for an FM with zero magnetization and is not involved in the shift of the hysteresis loop. Moreover, calculations according to a scheme described below show that modification of the distribution function (4) in the case of $I \lesssim J$ also hardly influences the other properties of the hysteresis loop (width, vertical shift, shape, etc.). For this reason, below we will assume for the sake of simplicity that $I = 0$, so that separation of the near-wall spins into a special group only implies that the distribution function (4) has to be multiplied by the factor $(1 - f)^4$. In addition, it is necessary to take into consideration the spin configurations near the domain wall, in which the energy of one magnetic bond in significantly enhanced (under favorable conditions) due to the aforementioned mechanism. These additional configurations and their probabilities are presented in Table 2, where symbols "⥣" and "⥥" indicate the near-wall bonds with increased (J_E) and "normal" (J) energies.

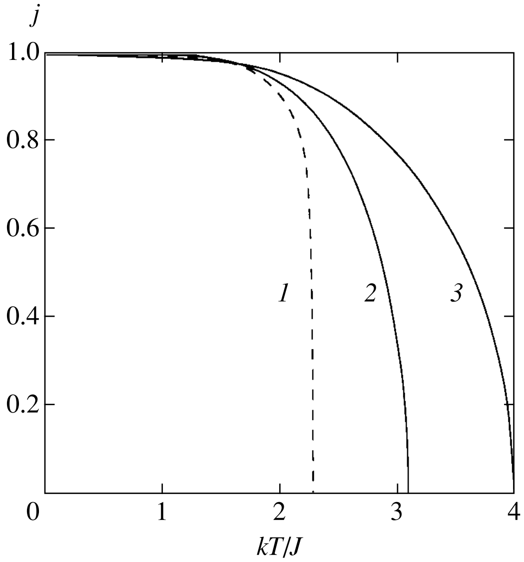


Fig. 2. The temperature dependences of the magnetization j calculated according to (*1*) the exact solution of the Ising problem on a square lattice, (*2*) the proposed generalized theory, and (*3*) the traditional mean field theory.

For $f \ll 1$, the configurations with more than one enhanced bond are rare and can be ignored. Then, the distribution function of the fields created by configurations involving the near-wall spins is as follows:

$$\begin{aligned} F_f(j; H) &= \frac{1}{16}[(1+j)^3\delta(h_E + 3h_J) \\ &+ 3(1+j)^2(1-j)\delta(h_E + h_J) + 3(1+j)(1-j)^2 \\ &\times \delta(h_E - h_J) + (1-j)^3\delta(h_E - 3h_J)] \\ &+ \frac{1}{16}[(1+j)^3\delta(2h_J) + 3(1+j)^2(1-j)\delta(0) \\ &+ 3(1+j)(1-j)^2\delta(-2h_J) + (1-j)^3\delta(-4h_J)], \end{aligned} \tag{8}$$

where $h_E = J_E/\mu = (J_E/J)h_J$. The total distribution function of random magnetic fields can be written as

$$F(j; H) = \frac{(1-f)^4 F_0(j; H) + 4f(1-f)^3 F_f(j; H)}{(1-f)^4 + 4f(1-f)^3}. \tag{9}$$

In this case, the magnetization $j(T)$ of the system predicted by Eq. (2) is nonzero only at an infinitely high temperature. This behavior is related to the assumption that the parameter f is independent of temperature. In real systems, where $f \longrightarrow 0$ as $T \longrightarrow T_N$, this drawback of the model is absent. However, in the case under consideration we have $T < T_N$ and, hence, the temperature dependence of the parameter f can be ignored.

Table 2. Additional configurations of the magnetic bonds and their probabilities on a square lattice

Configuration	Effective magnetic field H/h_J	Number of configurations	Probability of configuration (for magnetization j)
⇑↑↑↑	$J_E/J + 3$	$4C_3^0 = 4$	$4\left(\frac{f}{2}\right)(1-f)^3\left(\frac{1+j}{2}\right)^3$
⇑↑↑↓	$J_E/J + 1$	$4C_3^1 = 12$	$12\left(\frac{f}{2}\right)(1-f)^3\left(\frac{1+j}{2}\right)^2\left(\frac{1-j}{2}\right)$
⇑↑↓↓	$J_E/J - 1$	$4C_3^2 = 12$	$12\left(\frac{f}{2}\right)(1-f)^3\left(\frac{1+j}{2}\right)\left(\frac{1-j}{2}\right)^2$
⇑↓↓↓	$J_E/J - 3$	$4C_3^3 = 4$	$4\left(\frac{f}{2}\right)(1-f)^3\left(\frac{1-j}{2}\right)^3$
⇓↑↑↑	$+2$	$4C_3^0 = 4$	$4\left(\frac{f}{2}\right)(1-f)^3\left(\frac{1+j}{2}\right)^3$
⇓↑↑↓	0	$4C_3^1 = 12$	$12\left(\frac{f}{2}\right)(1-f)^3\left(\frac{1+j}{2}\right)^2\left(\frac{1-j}{2}\right)$
⇓↑↓↓	-2	$4C_3^2 = 12$	$12\left(\frac{f}{2}\right)(1-f)^3\left(\frac{1+j}{2}\right)\left(\frac{1-j}{2}\right)^2$
⇓↓↓↓	-4	$4C_3^3 = 4$	$4\left(\frac{f}{2}\right)(1-f)^3\left(\frac{1-j}{2}\right)^3$

The distribution (9) is asymmetric with respect to the magnetic field. For a system with zero magnetization ($j = 0$), the center-of-gravity of this function for $f \ll 1$ corresponds to the field

$$\bar{H}(0) = \int_{-\infty}^{\infty} HF(0; H)dH$$

$$\approx 4f\int_{-\infty}^{\infty} HF_f(0; H)dH = 2f(h_E - h_J).$$

If $J_E > J$, we have $\bar{H}(0) > 0$ and the hysteresis loop is shifted toward negative fields, otherwise it is shifted in the positive direction. We can expect that the center of the hysteresis loop corresponds to the field

$$H_{\rm EB} \sim -\bar{H}(0) \approx -2f(J_E/J - 1)h_J. \qquad (10)$$

However, this conclusion is valid (see below) only for $J_E \lesssim kT$. For large J_E, the function $H_{\rm EB}(J_E)$ exhibits saturation so that $H_{\rm EB} \longrightarrow (H_{\rm EB})_{\max} \sim 4fh_J$.

3. MAGNETIC MOMENT RELAXATION EQUATION

Consider the system of magnetic moments μ forming a 2D Ising spin lattice described by the Ising variables $s_j = \pm 1$ ($j = 1, 2, \ldots, N$). Thermal fluctuations at a finite temperature T lead to the spin flip from "↑" to "↓" state and back. The probabilities of such events determine the relaxation equations for the system state parameters. In particular, the probability $P(s_1, s_2, \ldots, +s_j, \ldots, s_N; t)$ of the realization of a given spin configuration at the time moment t is determined by the equation [9]

$$\begin{aligned}&\frac{d}{dt}P(s_1, s_2, \ldots, +s_j, \ldots, s_N; t)\\ &= -\sum_j w_j(s_j)P(s_1, s_2, \ldots, +s_j, \ldots, s_N; t)\\ &+ \sum_j w_j(-s_j)P(s_1, s_2, \ldots, -s_j, \ldots, s_N; t),\end{aligned} \qquad (11)$$

where $w_j(s_j)$ and $w_j(-s_j)$ are the probabilities of jth spin flip from the "↑" and "↓" states, respectively. According to the principle of detailed balancing, we have

$$\begin{aligned}&w_j(s_j)P_0(s_1, s_2, \ldots, +s_j, \ldots, s_N)\\ &= w_j(-s_j)P_0(s_1, s_2, \ldots, -s_j, \ldots, s_N),\end{aligned} \qquad (12)$$

where $P_0(s_1, \ldots)$ are the probabilities corresponding to the thermodynamically equilibrium configuration. These probabilities obey the Boltzmann relation

$$\frac{P_0(s_1, s_2, \ldots, -s_j, \ldots, s_N)}{P_0(s_1, s_2, \ldots, +s_j, \ldots, s_N)} = \exp\left(-\frac{\Delta_j E}{kT}\right), \qquad (13)$$

where, $\Delta_j E$ is the difference between the energies of states with opposite directions of the jth spin. Evidently, $\Delta_j E = 2E_j$, where

$$E_j = \mu H_e + \sum_{k \neq j} s_k J_{jk} \tag{14}$$

is the energy of the jth spin in the state "↑" (determined by its interaction with the external magnetic field H_e and the other spins) and J_{jk} is the energy of interaction between the jth and kth spins. Equations (12)–(14) yield

$$\frac{w_j(s_j)}{w_j(-s_j)} = \frac{1 - s_j \tanh(E_j/kT)}{1 + s_j \tanh(E_j/kT)}. \tag{15}$$

According to the Glauber approach [10], the probabilities $w_j(s_j)$ and $w_j(-s_j)$ obeying relation (15) are written as

$$\begin{aligned} w_j(s_j) &= \frac{1}{2}\Omega\left(1 - s_j \tanh\frac{E_j}{kT}\right), \\ w_j(-s_j) &= \frac{1}{2}\Omega\left(1 + s_j \tanh\frac{E_j}{kT}\right), \end{aligned} \tag{16}$$

where Ω is the parameter having the sense of a frequency at which spins try to change their orientations. In this approach, the equilibrium state of the system coincides with that according to the mean field theory (see below).

The configuration-average value of the jth spin at the time t is

$$\langle s_j \rangle = \sum_{(s)} s_j P(s_1, \ldots, s_N; t), \tag{17}$$

where the sum is taken over all spin configurations of the system. Substituting Eq. (11) into formula (17) and using expression (16), we obtain [9]

$$\frac{d}{dt}\langle s_j \rangle = -\Omega\left(\langle s_j \rangle - \left\langle \tanh\frac{E_j}{kT} \right\rangle\right). \tag{18}$$

In accordance with the ideology of the mean field theory, this equation can in principle be simplified by substituting

$$\left\langle \tanh\frac{E_j}{kT} \right\rangle \longrightarrow \tanh\frac{\langle E_j \rangle}{kT},$$

where

$$\langle E_j \rangle = \mu H_e + \sum_{k \neq j} J_{jk} \langle s_k \rangle.$$

Taking into account the fact that $j \propto \langle s_j \rangle$, we obtain an equation

$$\frac{dj}{dt} = -\Omega\left(j - \tanh\frac{\mu H_e + j \sum_{k \neq j} J_j k}{kT}\right) \tag{19}$$

describing relaxation of the magnetization toward the thermodynamically equilibrium state determined by the standard mean field relation:

$$j = \tanh\left[\left(\mu H_e + j \sum_{k \neq j} J_j k\right) \Big/ kT\right].$$

However, more exact results can be obtained, as was demonstrated above, by means of the generalization of the mean field model. According to this approach, Eq. (18) can be rewritten as

$$\frac{dj}{d(\Omega t)} = -j + \int_{-\infty}^{\infty} \tanh\left[\frac{\mu(H_e + H)}{kT}\right] F(j, H) dH. \tag{20}$$

This equation, together with the local field distribution function (9), determines the relaxation phenomena in the system under consideration and, in particular, describes the hysteresis loop.

4. HYSTERESIS LOOP

Prior to proceeding with an analysis of the relaxation equation (20), let us formulate some conclusions following from the form of this equation for the particular system under consideration.

(i) The hysteresis, as the relaxation phenomenon, significantly depends on the rate of magnetic field variation in the course of measurements. The natural time scale in this case is provided by the inverse frequency Ω. According to this, a process with the characteristic time $\tau \ll \Omega^{-1}$ is fast, while that with $\tau \gg \Omega^{-1}$ is slow. The typical process of measurement of the hysteresis loop is a slow process. However, even in this case the shape of the loop significantly depends on the period τ of variation of the external magnetic field and the law of its variation. This is illustrated in Fig. 3, which shows the hysteresis loops for various periods (Fig. 3a) and different laws of variation (Fig. 3b) of the magnetic field. As can be seen from this figure, the shape of the loop and especially its width (even for $\Omega\tau \gg 1$) strongly depend on the field sweep time. On the other hand, the law of the field variation (at $\Omega\tau = \text{const}$) rather insignificantly influences the loop width and virtually does not affect the shift. Below we describe the results obtained for a cosine-shaped field sweep (see the inset to Fig. 3b) with a total duration of $\tau = 320/\Omega$ at a temperature of $\Theta \equiv kT/J = 2$ and the magnetic parameters $h_J \equiv J/\mu = 1$ and $J_E = 10J$.

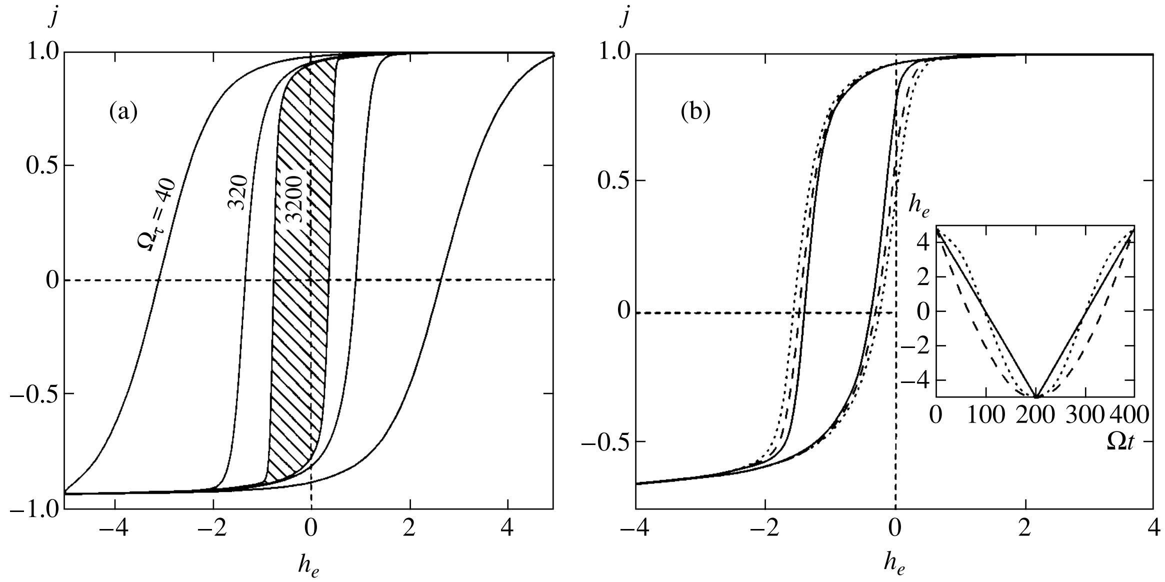


Fig. 3. The hysteresis loops calculated for (a) various periods and (b) different laws of variation of the magnetic field at $\Theta = 2$, $h_J = 1$, $J_E = 10J$, and $f = 0.01$ (a) and 0.05 (b). The inset shows the triangle (solid), sine (dash), and cosine (dot) shapes of the field sweep.

(ii) The shift of the loop along the field axis is caused, as was indicated above, by the asymmetric part $F_f(h_J)$ of the distribution function (8), which is centered at the field $\bar{H}(0) = 2f(J_E/J - 1)h_J$. This field exhibits a linear increase with the exchange energy J_E. However, this dependence does not imply an infinite leftward shift of the loop, since, in fact, the displacement is determined by the term

$$4f\int_{-\infty}^{\infty}\tanh\left[\frac{\mu(H_e+H)}{kT}\right]F_f(j, H)dH$$

in the right-hand part of Eq. (20). For $J_E \longrightarrow \infty$, this term tends to $4fh_J$ (in fact, this value is approached at $J_E/J \gtrsim \Theta$). Therefore, the dependence $H_{EB}(J_E)$ is expected to saturate as $J_E \longrightarrow \infty$.

A physical reason for this saturation is quite clear. For sufficiently low temperatures ($\Theta \lesssim 1$) and small amplitudes ($h_{e0} \lesssim J_E/J$) of the magnetic field sweep, only the FM spins with a weak coupling J respond to the field, while spins with enhanced coupling J_E retain their directions. In this case, spins of the first group create a symmetric hysteresis loop, while the second group accounts for its displacement leftward by the value proportional to the total magnetization $j_{EB} \sim f$ in this group. As a result, the center of the hysteresis loop exhibits a leftward displacement by $|H_{EB}|_{max}/h_J \approx j_{EB}/j_0'$, where $j_0' = dj/dh_e|_{f=0, j=0}$ is the slope of the symmetric (i.e., that for $f = 0$) hysteresis loop at the points $j = 0$ (i.e., at $H = \pm H_c$, where H_c is the coercive field for the symmetric loop). The parameter j_0' is determined from Eq. (20) for $f = 0$ and $j = 0$. In this case, the right-hand part of this equation is equal with a good approximation to $(3/16)h_e$. Then, for a linear field sweep with a period τ and an amplitude h_{e0}, we obtain $h_e(t) = h_{e0}(1 - 4t/\tau)$, and Eq. (20) yields $j_0' \sim (1/10)\Omega\tau/h_{e0}$. Therefore, the maximum displacement of the hysteresis loop can be estimated as

$$\frac{|H_{EB}|_{max}}{h_J} \sim \frac{1}{10}\frac{f\Omega\tau}{h_{e0}}.$$

For $\Omega\tau \sim 100$ and $h_{e0} \sim 1$, this yields $|H_{EB}|_{max}/h_J \sim 10f$. A comparison of this value to expression (10) shows that saturation of the dependence $|H_{EB}|_{max}(J_E/J)$ must take place at $J_E/J \sim 5$. Apparently, the loop width ceases to change as well.

Now let us proceed to an analysis of the hysteresis loop of the system under consideration, which is described by a solution to Eq. (20). We assume that the system is completely magnetized ($j = 1$) in the initial state ($h_e \equiv \mu H_e/J = 5$), after which the field is first decreased to $h_e = -5$ and then increased to the initial level.

Figure 4 shows the typical hysteresis loops for the system with various fractions f of enhanced bonds. As can be seen, an increase in f leads to a shift of the hysteresis loop toward negative fields. Simultaneously, the loop exhibits narrowing and shifts up toward greater positive magnetization. The first of these effects is described by the exchange bias parameter $H_{EB} = (H_c^- + H_c^+)/2$, where H_c^- and H_c^+ are the coercive fields corresponding to decrease and increase in the field. The lat-

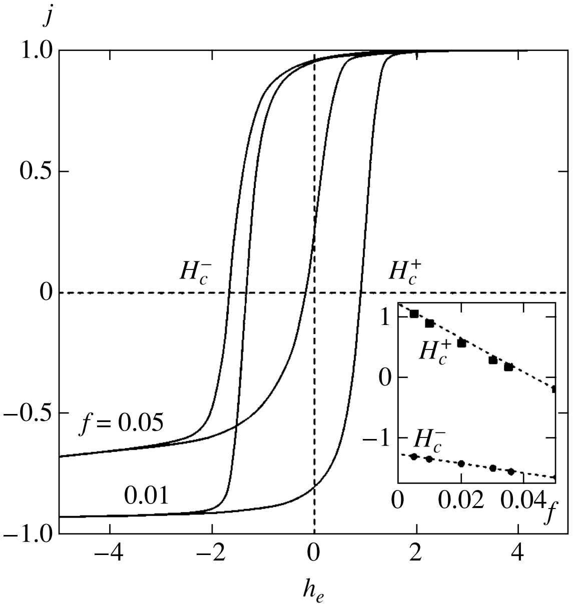


Fig. 4. The hysteresis loops calculated for various fractions f of enhanced bonds. The inset shows the plot of coercive fields versus f.

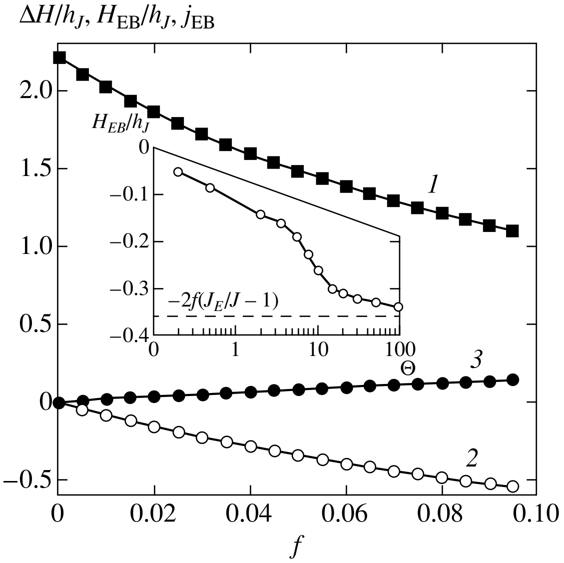


Fig. 5. Plots of the (*1*) hysteresis loop width ΔH, (*2*) exchange bias H_{EB}, and (*3*) magnetization increment j_{EB} versus fraction f of enhanced magnetic bonds. The inset shows the temperature dependence of the exchange bias ($f = 0.02$, $J_E/J = 10$).

ter trends are described by formulas for the loop width, $\Delta H = (H_c^+ + H_c^-)/2$, and the magnetization increment $j_{\mathrm{EB}} = (j_s^+ + j_s^-)/2$, where j_s^+ and j_s^- are the maximum and minimum magnetization values.

Figure 5 shows the dependences of all three parameters (ΔH, H_{EB}, and j_{EB}) on the fraction f of exchange-enhanced bonds. As can be seen, the shifts H_{EB} and j_{EB} of the hysteresis loop exhibit a linear increase, while the loop width ΔH shows a linear decrease with increasing f. The increase in H_{EB} and j_{EB} with f is quite understandable, since it reflects the growth in the number of enhanced bonds, which favor to provide positive magnetization of the system even in the region of negative fields. It should be noted that the behavior of $H_{\mathrm{EB}}(f)$ quantitatively disagrees with relation (10), which is explained by the aforementioned effect of the H_{EB} saturation at large J_E values. As for the narrowing of the hysteresis loop, this is related (see Fig. 4) for the most part to a significant shift of the coercive field H_c^+ to the negative values, whereas the coercive field H_c^- changes to a much lesser extent. The greater shift of H_c^+ is explained by the aforementioned tendency of the system to retain positive magnetization of the system even in the region of negative fields, which is caused by the presence of spins with enhanced coupling.

The inset to Fig. 5 shows the temperature dependence of the exchange bias $H_{\mathrm{EB}}(\Theta)$ of the hysteresis loop. The increase in the absolute EB value with the temperature is easy to understand: at higher temperatures, the system always occurs in an almost stationary state corresponding to the maximum possible value of H_{EB} determined by relation (10). As can be seen from Fig. 5, H_{EB} tends to this very level at high temperatures.

Figure 6 shows saturation of the H_{EB} and ΔH values with increasing energy J_E of enhanced coupling. This

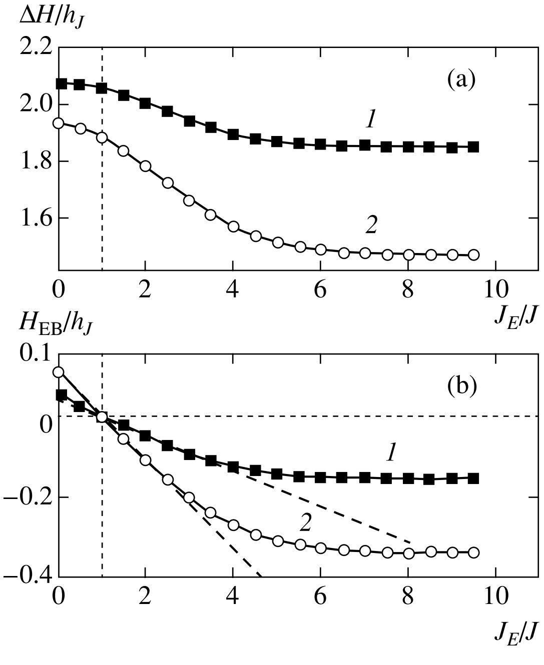


Fig. 6. Plots of the (a) hysteresis loop width ΔH and (b) exchange bias H_{EB} versus energy J_E of enhanced coupling for ($f = 0.02$ (*1*) and 0.05 (*2*). The dashed lines correspond to formula (10).

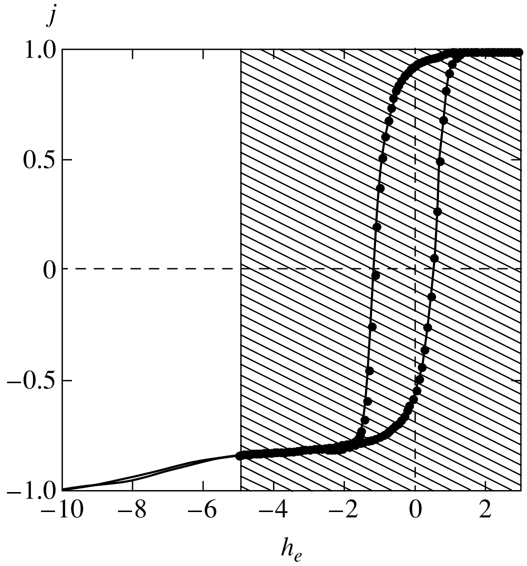


Fig. 7. The hysteresis loops calculated for the same field sweep rates dh_e/dt but various sweep amplitudes: $h_{\max} = 30$, $\Omega\tau = 2000$ (solid curve); $h_{\max} = 5$, $\Omega\tau = 333$ (points). Dashed region corresponds to $h_e > -h_{\max}$, where the two loops coincide.

behavior was discussed above and quantitatively agrees with the obtained conclusions.

The vertical shift of the hysteresis loop described by the parameter j_{EB} is related to a difference in the magnetization of the system for the positive and negative directions of the magnetic field (Fig. 4). Evidently, this difference must disappear for a sufficiently large absolute value of the negative field $H_e \gtrsim \mu J_E$. In our case, this corresponds to $h_e \gtrsim 10$ and agrees well with the results of calculations presented in Fig. 7.

An important and, hence, frequently measured characteristic of the real systems is the dependence of the exchange bias on the AFM film thickness. Although we have considered a 2D model, it is possible to qualitatively judge the shape of this dependence. Indeed, the fraction f of enhanced bonds is inversely proportional to the average size L of the AFM domain: $f \propto 1/L$. Since the formation and stabilization of domains in the AFM is caused by defects, the domain walls appear as closed surfaces (2D boundaries for the films of finite thickness) containing the maximum possible number of defects. Obviously, the possible number of such surfaces increases with the film thickness d. Since the minimum energy corresponds to the minimum possible (for the given conditions) length of this boundary, the average domain size L must also increase with the film thickness. Accordingly, the parameter f and, hence, the exchange bias H_{EB} must decrease with increasing d. Such a dependence of the exchange bias on the AFM film thickness was frequently observed in experiment [1].

5. CONCLUSIONS

The main points of the proposed model, which make possible its analytical consideration, are the presence of a small fraction of enhanced (by any mechanism) ferromagnetic bonds, generalization of the mean field theory through introduction of the distribution function of the spin coupling energies, and the corresponding generalization of the magnetic relaxation energy.

Within the framework of the developed generalized model, we considered the exchange bias of the hysteresis loop of a 2D Ising FM–AFM bilayer, determined various characteristics of the hysteresis loop, and analyzed their dependence on the parameters of the model. On the whole, the obtained results agree with experiment.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research, project nos. 03-02-17029, 04-02-19964, and 05-02-17021.

REFERENCES

1. J. Nogués and I. K. Schuller, J. Magn. Magn. Mater. **192**, 203 (1999).
2. M. Kiwi, J. Magn. Magn. Mater. **234**, 584 (2001).
3. M. Lederman, IEEE Trans. Magn. **35**, 794 (1999).
4. U. Nowak, K. D. Usadel, J. Keller, *et al.*, Phys. Rev. B **66**, 14430 (2002).
5. X. Illa, E. Vives, and A. Planes, Phys. Rev. B **66**, 224422 (2002).
6. M. W. Klein and R. Brout, Phys. Rev. **132**, 2412 (1963).
7. M. Thomsen, M. F. Thorpe, T. C. Choy, and D. Sherrington, Phys. Rev. B **30**, 250 (1984); T. C. Choy, D. Sherrington, M. Thomsen, and M. F. Thorpe, Phys. Rev. B **31**, 7355 (1985); M. Thomsen, M. F. Thorpe, T. C. Choy, and D. Sherrington, Phys. Rev. B **33**, 1931 (1986).
8. R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic, New York, 1982; Mir, Moscow, 1985).
9. M. Suzuki and R. Kubo, J. Phys. Soc. Jpn. **24**, 51 (1968).
10. R. J. Glauber, J. Math. Phys. **4**, 294 (1963).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 100, No. 6, 2005, pp. 1121–1126.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 6, 2005, pp. 1272–1278.
Original Russian Text Copyright © 2005 by Troyanchuk, Karpinskiĭ, Dobryanskiĭ, Fedotova, Szymczak.

ORDER, DISORDER, AND PHASE TRANSITIONS IN CONDENSED SYSTEMS

The Two-Phase Crystal Structure and Magnetic Properties of the $LaCo_{1-x}Fe_xO_{3-d}$ System

I. O. Troyanchuk[a],*, **D. V. Karpinskiĭ**[a], **V. M. Dobryanskiĭ**[a],
Yu. A. Fedotova[b], **and H. Szymczak**[c]
[a]*Institute of Solid-State and Semiconductor Physics, Belarussian Academy of Sciences,*
ul. Brovki 17, Minsk, 220072 Belarus
[b]*National Scientific-Teaching Center of Physics of Particles and High Energies, Minsk, 220040 Belarus*
[c]*Institute of Physics, Academy of Sciences of Poland, 02-668 Warsaw, Poland*
**e-mail: troyan@ifttp.bas-net.by*

Received October 25, 2004

Abstract—The preparation conditions and magnetic properties of the $La(Co_{1-x}Fe_x)O_{3-d}$ system of solid solutions were studied. Irrespective of synthesis conditions, the samples with $x \leq 0.15$ were rhombohedral (space group $R\bar{3}c$), and those with $x \geq 0.7$, orthorhombic (space group *Pnma*). In the intermediate region of $0.15 < x < 0.7$, the samples could be mixtures of rhombohedral and orthorhombic phases. Structural heterogeneity considerably decreased as the temperature of the synthesis increased or after annealing in a reducing medium. The samples containing more than 40% iron exhibited weak spontaneous magnetization, as is characteristic of weak ferromagnets. The magnetic properties and structure depend on the content of oxygen. A decrease in the concentration of oxygen sharply increases magnetic anisotropy; the coercive force is then close to 10 kOe, which is more than that known for other oxide systems. The magnetic properties are explained in terms of the model according to which Co^{3+} ions are in the low-spin state. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Cobaltites with perovskite structures exhibit several unusual phase transitions related to changes in the spin state of ions [1–5]. The trivalent cobalt ions (Co^{3+}) in $LaCoO_3$ experience the transition from the low-spin (LS) nonmagnetic ground state with $S = 0$ to the intermediate spin (IS) state with $S = 1$ at about $T = 90$ K. Close to $T = 500$ K, a metal-dielectric transition extended along the temperature axis is observed [1–4]. The specific resistance ρ obeys the law $\rho(T) \propto \exp(E_g/k_BT)$ at low temperatures; the ρ value decreases by several orders of magnitude at $T = 500$ K and increases as the temperature rises at the higher temperatures [1]. The spin state transition at $T = 90$ K disappears for the $La_{1-x}Sr_xCoO_3$ system doped with Sr, and the cobalt ions then remain magnetic even to the lowest temperature. Systems with $x > 0.2$ exhibit spontaneous magnetization at low temperatures [6]. The magnetization and neutron diffraction data led several authors to suggest the existence of the state of a cluster spin glass and ferromagnetism in the region $0.3 < x < 0.5$ [7]. At lower strontium contents ($x < 0.2$), magnetization measurements revealed the spin-glass ground state [7, 8] with strong ferromagnetic short-range order correlations according to the experimental diffuse neutron scattering data [8]. The suggestion was made that lattice expansion caused by the substitution of Sr stabilized the intermediate spin state of Co^{3+} [8], whereas the appearance of ferromagnetic short-range order could be caused by the introduction of charge carriers. The latter factor should also change electron transport properties. The $\rho(T)$ value for $La_{1-x}Sr_xCoO_3$ with $x < 0.2$ is smaller by several orders of magnitude than for $LaCoO_3$, although $\rho(T)$ remains semiconducting in character [9]. According to the thermal electromotive force and Hall coefficient signs, the compound has hole conduction [9]. Above $T = 500$ K, $\rho(T)$ increases as the temperature grows, which presupposes the occurrence of the dielectric–metal transition at $T = 500$ K, as in $LaCoO_3$. Thermal electromotive force changes are also indicative of the occurrence of the dielectric–metal transition [9].

It was found in [10, 11] that the magnetic state of the Co^{3+} ions in $LaCoO_3$ can be changed not only by substituting Sr but also by replacing Co^{3+} with isovalent Ni^{3+} ions. At $x < 0.5$, $La(Co_{1-x}Ni_x)O_3$ samples exhibit spin-glass properties at low temperatures. In this work, we show that the replacement of Co^{3+} with Fe^{3+} does not cause significant stabilization of the intermediate spin state of cobalt ions. This system exhibits giant magnetic anisotropy at $x \geq 0.5$; the mechanism of its formation is discussed. Earlier, the $LaCo_{1-x}Fe_xO_3$ system was studied from the point of view of its catalytic activity [11].

1063-7761/05/10006-1121$26.00 © 2005 Pleiades Publishing, Inc.

2. EXPERIMENTAL

The $La(Co_{1-x}Fe_x)O_{3-d}$ samples were prepared in air following the usual ceramic technology from a mixture of La_2O_3, Co_3O_4, and Fe_2O_3 oxides at temperatures from $T = 1470$ K to $T = 1770$ K. Several synthesized samples were reduced in evacuated quartz ampules at $T = 1200$ K. Tantalum metal was used as a getter for oxygen. The X-ray measurements were performed on a DRON-3M diffractometer using CuK_α radiation. The crystal structure was calculated with the use of the FullProf program. The magnetic measurements were performed on an OI-1001 commercial vibrating-coil magnetometer and using an MPMS-5 (Quantum Design) SQUID magnetometer. The Mössbauer spectra were recorded on an MS2000 spectrometer in the transmission mode using a ^{57}Co/Rh source. The spectra were calculated with the help of the MOSMOD program. The isomeric shifts are given with respect to α-Fe. The topography of the surface was studied with a KARL ZEISS scanning electron microscope. X-ray spectral analysis data were obtained using a Röntec energy dispersion Si–Li semiconductor detector.

3. RESULTS AND DISCUSSION

Crystal structure. FullProf calculations for the samples prepared at $T = 1570$ K in air were successfully performed to show the samples with $x \le 0.15$ to be single-phase, space group $R\bar{3}c$; the samples with $0.3 \le x \le 0.6$ to be mixtures of rhombohedral and orthorhombic (*Pnma*) phases; and the samples with $x \ge 0.7$ to be single-phase orthorhombic (space group *Pnma*).

Increasing the synthesis temperature to $T = 1770$ K sharply decreased crystal structural inhomogeneity. The samples prepared at this temperature were successfully calculated as rhombohedral ($x \le 0.5$) and orthorhombic ($x \ge 0.6$). Only the sample with $x = 0.55$ could not be calculated in terms of a single-phase model. The unit cell volume increased almost linearly as the content of iron grew in the interval $0 \le x \le 1$. It is likely that high temperature contributes to the formation of homogeneous solid solutions. There is, however, another method for the preparation of samples with homogeneous crystal structures; it involves thermal treatment in a reducing medium. The FullProf calculation results obtained for the samples with $x = 0.5$ prepared at $T = 1570$ K and $T = 1770$ K in air and the sample reduced in an evacuated quartz ampule in the presence of tantalum metal at $T = 1170$ K are shown in Fig. 1. Reduction was performed for the sample prepared at $T = 1570$ K in air. Treatment of the sample in an evacuated ampule caused weight loss corresponding to a 2% decrease in the content of oxygen. The sample was fully transformed into the orthorhombic phase. The unit cell parameters of several samples are listed in the table. The unit cell volume increases as the content of oxygen decreases. This can be explained by the appearance of Co^{2+} ions, whose ionic radius is larger than that of Co^{3+}. Electron microscopic studies of the surface of samples showed the presence of 2–10 μm polycrystalline grains in the samples with $x = 0.5$. According to the X-ray spectral microanalysis data, the content of cobalt and iron in reduced sample grains was homogeneous. Nor did we observe significant deviations from a uniform distribution of cobalt and iron for the samples prepared in air.

Magnetic properties. Samples with low iron contents did not exhibit a dependence of magnetization on their magnetic prehistory, as is characteristic of paramagnets or antiferromagnets. Nor did we observe anomalous behavior indicative of antiferromagnetic ordering. The reciprocal susceptibilities measured for the samples with $x = 0.1$, 0.15, and 0.4 are shown in Fig. 2. The absolute value of the asymptotic paramagnetic Curie point increased as the content of iron grew. This is evidence of strengthening of antiferromagnetic exchange interactions.

The field dependences of magnetization obtained at various temperatures are close to linear, as is characteristic of paramagnets or antiferromagnets. The field dependence of magnetization for spin glasses should be substantially nonlinear.

Note that the magnetic properties of the sample doped with iron ions (10%) are sharply different from those of pure $LaCoO_3$, which experiences gradual transition into the diamagnetic state at temperatures below

Table

	Space group	Unit cell parameters, Å	Volume, Å^3
$LaCo_{0.6}Fe_{0.4}O_3$	$R\bar{3}c$	$a = 5.4784(7)$, $c = 13.2127(6)$	57.132
$LaCo_{0.5}Fe_{0.5}O_3$	$R\bar{3}c$	$a = 5.4881(7)$, $c = 13.2368(6)$	57.539
$LaCo_{0.5}Fe_{0.5}O_{3-d}$ reduced	*Pnma*	$a = 5.4627(7)$, $b = 7.7573(8)$, $c = 5.5129(1)$	58.405
$LaCo_{0.4}Fe_{0.6}O_3$	*Pnma*	$a = 5.4405(9)$, $b = 7.3355(6)$, $c = 5.4936(4)$	57.7865
$LaCo_{0.3}Fe_{0.7}O_3$	*Pnma*	$a = 5.4742(0)$, $b = 7.7484(8)$, $c = 5.5070(3)$	58.4265

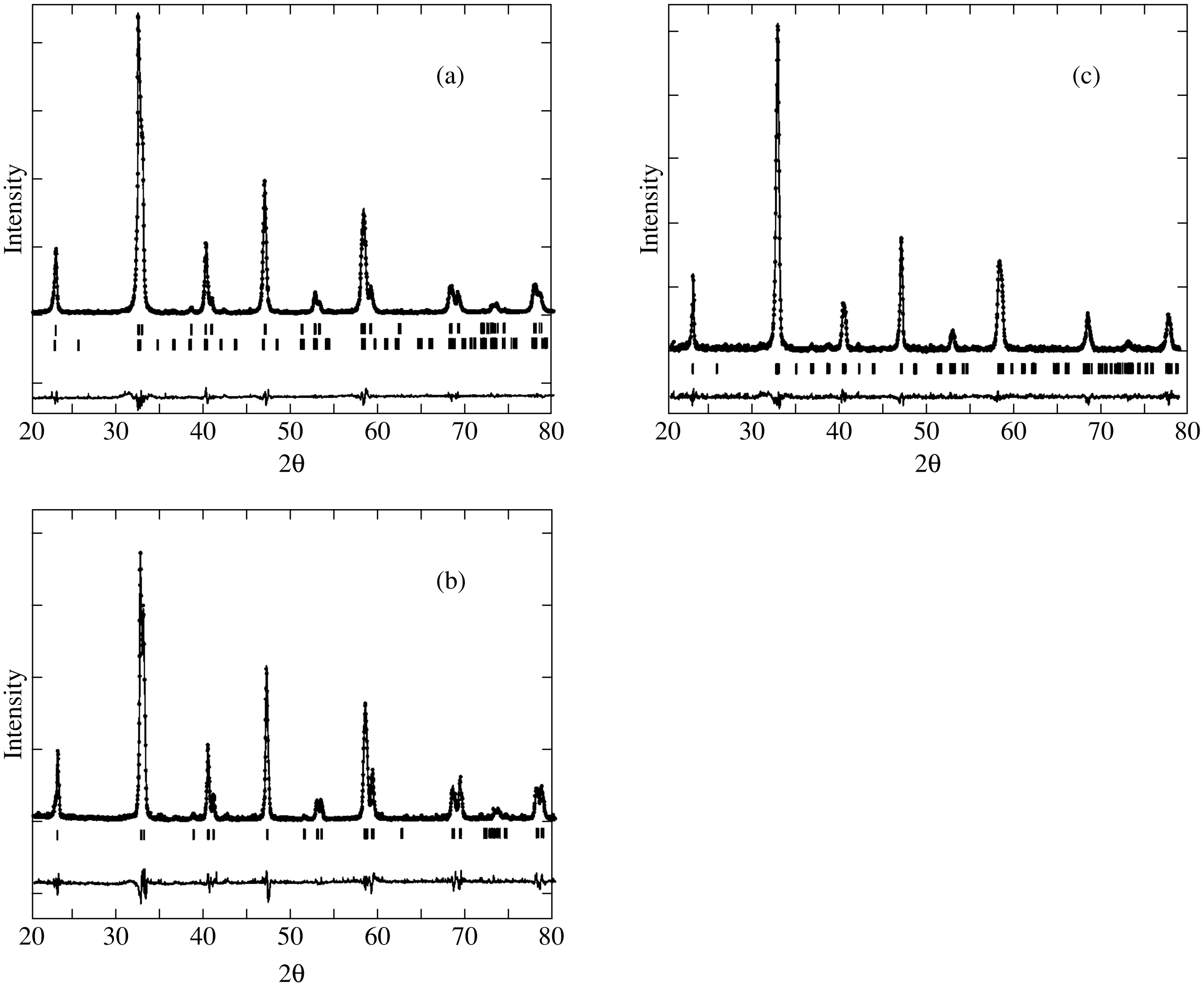


Fig. 1. Refinement of the structure of $La(Co_{0.5}Fe_{0.5})O_3$ prepared (a) at $T = 1570$ K in air, (b) at $T = 1770$ K in air, and (c) by reduction in a quartz ampule. The experimental data are given by dots, the calculated curve by a solid line, and their difference by the lower solid line; vertical dashes correspond to the calculated reflection positions. The upper row of dashes in Fig. 1a is for the rhombohedral phase, and the lower row, for the orthorhombic phase.

100 K [1]. According to magnetization measurements, this transition is absent in the sample with the concentration $x = 0.1$, and the sample remains paramagnetic even at $T = 5$ K. The properties of the sample with $x = 0.15$ did not differ qualitatively from those of the sample with $x = 0.1$.

All the samples rich in iron ($x \geq 0.4$) are characterized by a small spontaneous magnetic moment. The results obtained in studying the magnetic properties of the samples with $x = 0.4$ and $x = 0.5$ prepared at $T = 1770$ K and the remanent magnetization of the reduced composition are shown in Fig. 3. The sample with $x = 0.4$ becomes paramagnetic at $T_N = 120$ K. The transition is fairly sharp, which is not characteristic of spin glasses. The ZFC and FC curves at low temperatures show that magnetization increases as the temperature grows, which is also unusual for spin glasses. Possibly, the sample exhibits long-range magnetic order. However, conclusions about the type of its magnetic state are difficult to make in the absence of neutron diffraction data. The temperature of magnetic ordering sharply increases to 300 K as the content of iron grows to $x = 0.5$. The field dependences of magnetization are evidence that spontaneous magnetization is small, as is characteristic of weak ferromagnets. It was found for the sample with $x = 0.5$ prepared at 1570 K in air and reduced in a vacuum that it had a small spontaneous moment at room temperature; this moment could not be suppressed by heating the sample to 360 K (this was the highest temperature of our measurements). It follows that the critical temperature at which magnetic ordering disappears is higher than 360 K. Magnetization measurements for the samples heated after cooling to helium temperatures were performed without applying an external magnetic field. The reduced sample exhibited anomalous magnetization behavior over the temperature range 5–100 K. This behavior corresponded to

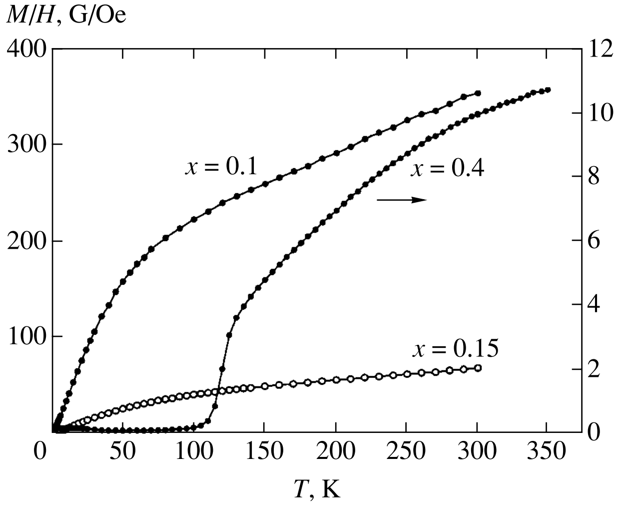


Fig. 2. Reciprocal paramagnetic susceptibility as a function of temperature for $La(Co_{0.9}Fe_{0.1})O_3$, $La(Co_{0.85}Fe_{0.15})O_3$, and $La(Co_{0.6}Fe_{0.4})O_3$.

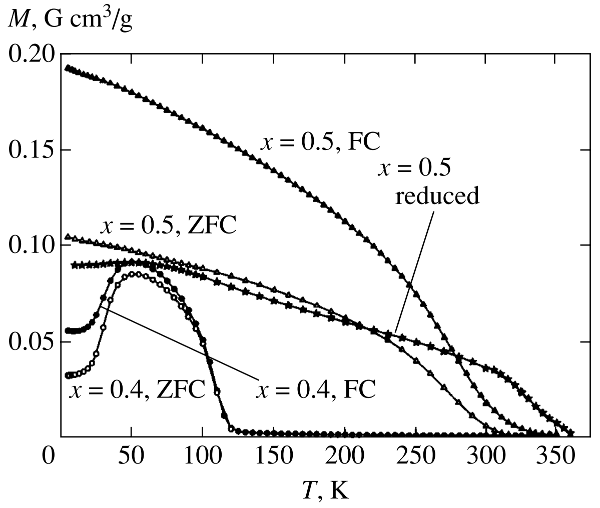


Fig. 3. Temperature dependences of FC and ZFC magnetizations for (○, ●) $La(Co_{0.6}Fe_{0.4})O_3$ and (△, ▲) $La(Co_{0.5}Fe_{0.5})O_3$ in a 100 Oe field. Shown by asterisks is the temperature dependence of the remanent magnetization of the reduced $La(Co_{0.5}Fe_{0.5})O_{3-d}$ sample.

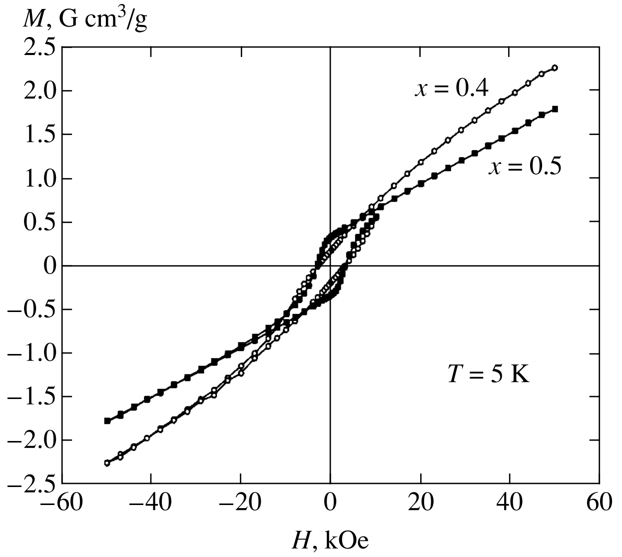


Fig. 4. Field dependences of magnetization for $LaCo_{0.5}Fe_{0.5}O_3$ and $LaCo_{0.6}Fe_{0.4}O_3$ at $T = 5$ K.

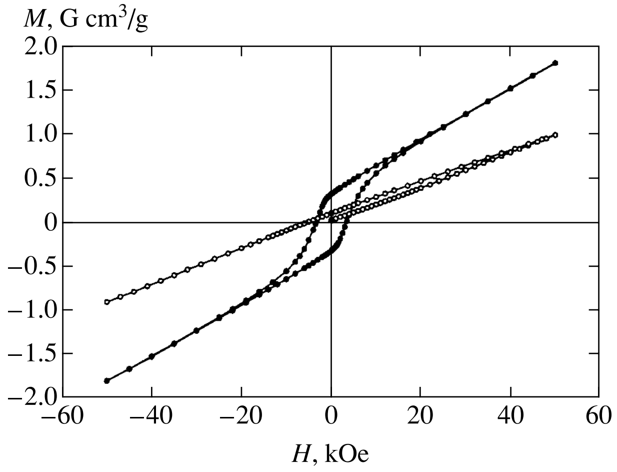


Fig. 5. Field dependences of magnetization for the reduced $LaCo_{0.5}Fe_{0.5}O_{3-d}$ sample at (●) $T = 293$ K and (○) $T = 10$ K.

a phase transition, possibly caused by spin reorientation.

Field dependences of magnetization show that the reduced composition with $x = 0.5$ was characterized by strong magnetic anisotropy. Hysteresis loops for the samples with $x = 0.4$ and $x = 0.5$ obtained in air at 1770 K are shown in Fig. 4. These samples are characterized by magnetic rigidity and a coercive force of about 3 kOe at $T = 5$ K. Oxygen loss during reduction causes a dramatic increase in magnetic anisotropy (Fig. 5). The coercive force is then about 10 kOe at $T = 293$ K. Such a magnetic anisotropy value at room temperature is unique for oxide compounds. Magnetic anisotropy sharply increases as the temperature decreases. For the reduced sample, a 50 kOe external magnetic field was too low to cause reorientation of magnetic moments at 10 K.

Conductivity measurements showed that the compositions with $0.3 \leq x \leq 0.5$ were characterized by the semiconducting conductivity type up to 370 K.

Mössbauer spectra. The Mössbauer spectrum of the sample with $x = 0.5$ prepared at 1770 K (rhombohedral phase) was recorded at $T = 293$ K. It was a doublet with quadrupole splitting, $\Delta Q = 0.440$ mm/s. The isomeric shift $\delta = 0.369$ mm/s was evidence that the iron ions were in the trivalent state. The spectra of orthorhombic samples consisted of two or more sextets, which was evidence of magnetic nonequivalence of the iron ions.

Discussion. As follows from the Mössbauer data, the iron ions in the $LaCo_{1-x}Fe_xO_3$ system are in the oxidation state 3+. The question arises what is the spin state of the cobalt ions? To answer this question, let us compare the properties of $LaCo_{1-x}Fe_xO_3$ and $LaCo_{1-x}Ni_xO_3$. The authors of [12] performed a complex study of the latter system to find that the cobalt and nickel ions were also in the trivalent state. The unit cell volume of the $LaCo_{0.5}Ni_{0.5}O_3$ composition was substantially larger than the unit cell volumes for the extreme solid solution compositions $LaCoO_3$ and $LaNiO_3$. This means that the cobalt ions in $LaCo_{1-x}Ni_xO_3$ were in the intermediate spin state, because the ionic radius of Co^{3+} (IS) is substantially larger than the ionic radius of Co^{3+} (LS) [13].

We, however, found that the unit cell volume of $LaCo_{1-x}Fe_xO_3$ solid solutions with homogeneous crystal structures linearly increased as the content of iron grew, which was evidence in favor of the low-spin state of the cobalt ions, as in $LaCoO_3$. Magnetic properties also show that the cobalt ions are in the low-spin state. For instance, let us compare the magnetic properties of $LaCo_{0.9}Ni_{0.1}O_3$ and $LaCo_{0.9}Fe_{0.1}O_3$. The former is a spin glass with T_f = 16 K, whereas the latter remains paramagnetic even at liquid helium temperatures. This shows that the cobalt ions are in the ground nonmagnetic state with the $t_{2g}^6e_g^0$ configuration. A sharp increase in magnetic susceptibility at low temperatures and as the concentration of iron ions increases is in all probability caused by the magnetic contribution of Fe^{3+} ions, whose exchange interaction is antiferromagnetic. These ions begin to form clusters as their concentration increases, and long-range magnetic order arises close to the $x = 0.4$ concentration. The magnetic properties of the composition with $x = 0.4$ more closely correspond to a magnetically ordered than spin glass state. This follows from a fairly sharp transition to the paramagnetic state, which is characteristic of systems with cooperative magnetic ordering. Neutron diffraction data are necessary to refine the magnetic structure of the compound. The temperature of the transition into the paramagnetic state sharply increases at $x = 0.5$, because the cobalt (LS) ions are nonmagnetic. The reason for this may be very strong exchange interaction between the iron ions. Indeed, the temperature of antiferromagnetic ordering in $LaFeO_3$ is 750 K [14], which is evidence of strong negative exchange interactions. The small spontaneous magnetization is in all probability caused by antisymmetric exchange interaction of the Dzyaloshinski–Moriya type. Indeed, both rhombohedral ($R\bar{3}c$) and orthorhombic (*Pbnm*) phases admit the appearance of weak ferromagnetism.

The compounds with $x = 0.4$ and $x = 0.5$ are magnetically rigid materials with a fairly high coercive force. Like Co^{3+} (LS), Fe^{3+} ions cannot make a considerable contribution to magnetic anisotropy. Nor can the large magnetic anisotropy value be caused by the morphology of crystallites, because, according to the electron microscopic data, the mean size of crystallites is about 10 μm. A crystallite should therefore be divided into a large number of magnetic domains. In all probability, the large magnetic anisotropy is caused by the Dzyaloshinski–Moriya antisymmetric exchange directly related to spin-orbit coupling. The reduction of the sample with $x = 0.5$ sharply increases magnetic anisotropy. At helium temperatures, a 5 T external field is insufficient for displacing the magnetic moment from the easy magnetic axis in the direction of the field. We believe that the reduction of samples causes the appearance of oxygen vacancies. Near these vacancies, Co^{2+} or Co^{3+} ions in the high-spin state are stabilized. Both of them can make a large contribution to single-ion magnetic anisotropy and to the anisotropy of weak ferromagnetism because of strong spin-orbit coupling [15]. The appearance of Co^{2+} (or Co^{3+} (HS)) magnetic ions somewhat increases Neél temperature, most likely because of negative exchange interactions between these ions and Fe^{3+}.

Note in conclusion that, in spite of a noticeable increase in the Co–O mean distance in CoO_6 octahedra as cobalt ions are replaced by iron, cobalt ions predominantly remain in the low-spin state. This is evidence that Co^{3+} (LS) remains in the ground magnetic state in $LaCoO_3$ as the temperature increases and the anomalous behavior of the properties of $LaCoO_3$ is caused by thermal excitations in Co^{3+} (IS) states. The magnetic properties of the $La(Co_{1-x}Fe_x)O_3$ system can be described using the model according to which Co^{3+} ions do not participate actively in exchange interactions, and weak ferromagnetism and large magnetic anisotropy are caused by the magnetically active subsystem of Fe^{3+} ions. Their interactions result in the appearance of long-range magnetic order at $x \approx 0.4$.

ACKNOWLEDGMENTS

This work was financially supported by the Belarussian Foundation for Basic Research (project nos. F03-155 and F03-120) and the State Committee for Scientific Research (Poland), grant no. KBN 1 P03B 038 27.

REFERENCES

1. R. R. Heikes, R. C. Miller, and R. Mazelsky, Physica (Amsterdam) **30**, 1600 (1964).
2. G. H. Jonker, J. Appl. Phys. **37**, 1424 (1966).
3. K. Asai, O. Yokokura, M. Suzuki, *et al.*, J. Phys. Soc. Jpn. **66**, 967 (1997).

4. K. Asai, A. Atsuro, O. Yokokura, *et al.*, J. Phys. Soc. Jpn. **67**, 290 (1998).
5. I. O. Troyanchuk, N. V. Kasper, D. D. Khalyavin, *et al.*, Phys. Rev. B **58**, 2418 (1998).
6. P. M. Raccah and J. B. Goodenough, J. Appl. Phys. **39**, 1209 (1968).
7. M. Itoh, I. Natori, S. Kubota, *et al.*, J. Phys. Soc. Jpn. **63**, 1486 (1994).
8. K. Asai, O. Yokokura, N. Nishimori, *et al.*, Phys. Rev. B **50**, 3025 (1994).
9. V. P. Gerthsen and K. H. Härdil, Z. Naturforsch. A **17**, 514 (1962).
10. Y. Kobayashi, S. Murata, K. Asai, *et al.*, J. Phys. Soc. Jpn. **68**, 1011 (1999).
11. L. Bedel, A. C. Roger, C. Estournes, *et al.*, Catal. Today **85**, 207 (2003).
12. T. Kyômen, R. Yamazaki, and M. Itoh, Phys. Rev. B **68**, 104416 (2003).
13. P. G. Radaelli and S.-W. Cheong, Phys. Rev. B **66**, 094408 (2002).
14. W. C. Koehler and E. O. Wollan, J. Phys. Chem. Solids **2**, 100 (1957).
15. K. P. Belov, *Rare-Earth Magnetic Substances and Their Application* (Nauka, Moscow, 1980) [in Russian].

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 100, No. 6, 2005, pp. 1127–1128.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 6, 2005, pp. 1279–1281.
Original English Text Copyright © 2005 by Vekilov, Salikhov, Isaev, Johansson.

ORDER, DISORDER, AND PHASE TRANSITIONS IN CONDENSED SYSTEMS

Ferromagnetism in Quasicrystals: Symmetry Aspects[¶]

Yu. Kh. Vekilov[a], S. V. Salikhov[a], E. I. Isaev[a], and B. Johansson[b,c]
[a]*Moscow Steel and Alloys Institute, Moscow, 119049 Russia*
[b]*Royal Institute of Technology (KTH) SE-10044, Stockholm, Sweden*
[c]*Uppsala University Box 530, S-751121, Uppsala, Sweden*
e-mail: yuri_vekilov@yahoo.com

Received November 5, 2004

Abstract—Magnetic-group analysis of the symmetries typical of quasicrystals shows that ferromagnetism is incompatible with the icosahedral symmetry. Depending on the magnetic field direction, the icosahedral symmetry in the magnetic field is reduced to pentagonal, trigonal or rhombic symmetries. © *2005 Pleiades Publishing, Inc.*

Quasicrystals are currently attracting great interest because of the variety of their unusual physical properties. Among these, the magnetic properties and, especially, ferromagnetism are the least studied. The first experimental information on ferromagnetism in quasicrystals was connected with the presence of ferromagnetic nonquasicrystalline inclusions of a second phase [1, 2], and later ferromagnetic-like behavior was observed in pure icosahedral phases *i*-AlPdMnB and *i*-AlPdFeB [3, 4]. Mössbauer spectroscopy and nuclear magnetic resonance (NMR) experiments have shown that the magnetic state in these systems was ferromagnetic, but the samples consisted of large magnetic clusters with extensions of about 20 nm [2, 4]. Theoretically, the ferromagnetic ordering in quasicrystals has never been considered, and the most important problem is the compatibility of ferromagnetism with quasicrystalline symmetries. To the best of our knowledge, it has never been considered before, although the color groups for quasicrystalline solids have been studied [5, 6]. In the present work, a theoretical analysis of the magnetic groups for quasicrystals is performed to solve this problem.

The existence of a magnetic structure in solids is formally connected with the time inversion operation. The time inversion symmetry operation R changes the direction of the current density in a solid but does not act on spatial coordinates. The element R commutes with the rotations C_n, the rotations S_{2n} and the reflections σ, and at the same time, $R^2 = E$ (the identity transformation). Magnetic crystalline classes for periodic solids are described in [7]. Structurally, magnetic classes can be divided into three types. The magnetic class of type I has the ordinary point group symmetry. The direct product of point groups with the group $\{E, R\}$ forms the magnetic class of type II. The magnetic class of type III contains the operation R in combination with some rotations or reflections. These magnetic classes have the structure $\mathbf{G}(\mathbf{H}) = \mathbf{H} + \mathbf{R}g\mathbf{H}$, where $\mathbf{H}$ is an invariant subgroup of index 2, $g \in \mathbf{G}$, but $g \notin \mathbf{H}$, and $\mathbf{R}$ is the time inversion operator. In periodic solids, 58 magnetic classes of type III exist. In this paper, we obtain the magnetic classes for quasiperiodic structures following the method described in [7]

We first obtain the magnetic classes for symmetries with a preferable main axis (we have the five-, eight-, ten-, and twelvefold axes in mind). Classes with such a symmetry are related to pentagonal, octagonal, decagonal, and dodecagonal systems, correspondingly. The magnetic classes of type I represent an ordinary point group. Ferromagnetism is possible in the following classes of type I: $\mathbf{C}_5$, $\mathbf{C}_{5h}$, $\mathbf{C}_8$, $\mathbf{C}_{8h}$, $\mathbf{S}_8$, $\mathbf{C}_{10}$, $\mathbf{C}_{10h}$, $\mathbf{S}_{10}$, $\mathbf{C}_{12}$, $\mathbf{C}_{12h}$, and $\mathbf{S}_{12}$, with the magnetic moment vector directed along the main axis. Of course, ferromagnetism is impossible in all type-II magnetic classes (due to the presence of the time inversion R).

As pointed out above, all possible index-2 subgroups should be determined in order to find the magnetic classes of type III. The simplest way is to use the tables of characters of irreducible representations. Classes with characters equal to one in one-dimensional representations form invariant subgroups of index 2. All possible groups are given in the table for the systems under consideration. Each magnetic class is defined by a point group and its index-2 subgroup (in parentheses), which is given in the table. We note that only the class C_5 does not have index-2 subgroups, and consistently magnetic classes of type III. For the existence of the ferromagnetic state in magnetic classes of type III, it is necessary that these classes do not contain the elements RI or R_{σ_h} (I is the spatial inversion and σ_h is the reflection in the plane perpendicular to the main axis). All classes of type III that allow a ferromagnetic state are given in the last column in the table. The

---

[¶] This article was submitted by the authors in English.

Magnetic classes for quasicrystalline symmetries

Symmetry	Point groups	Magnetic classes of type III	Ferromagnetic classes
Pentagonal	$\mathbf{C}_5$, $\mathbf{C}_{5v}$, $\mathbf{D}_5$, $\mathbf{S}_{10}$, $\mathbf{D}_{5d}$	$\mathbf{C}_{5v}(\mathbf{C}_5)$, $\mathbf{D}_5(\mathbf{C}_5)$, $\mathbf{S}_{10}(\mathbf{C}_5)$, $\mathbf{D}_{5d}(\mathbf{C}_{5v}, \mathbf{D}_5, \mathbf{S}_{10})$	$\mathbf{C}_{5v}(\mathbf{C}_5)$, $\mathbf{D}_5(\mathbf{C}_5)$, $\mathbf{D}_{5d}(\mathbf{S}_{10})$
Octagonal	$\mathbf{C}_8$, $\mathbf{S}_8$, $\mathbf{C}_{8v}$, $\mathbf{D}_8$, $\mathbf{C}_{8h}$, $\mathbf{D}_{4d}$, $\mathbf{D}_{8h}$	$\mathbf{C}_8(\mathbf{C}_4)$, $\mathbf{S}_8(\mathbf{C}_4)$, $\mathbf{C}_{8v}(\mathbf{C}_8, \mathbf{C}_{4v})$, $\mathbf{D}_8(\mathbf{C}_8, \mathbf{D}_4)$, $\mathbf{C}_{8h}(\mathbf{C}_8, \mathbf{C}_{4h}, \mathbf{S}_8)$, $\mathbf{D}_{4d}(\mathbf{D}_4, \mathbf{C}_{4v}, \mathbf{S}_8)$, $\mathbf{D}_{8h}(\mathbf{C}_{8h}, \mathbf{C}_{8v}, \mathbf{D}_8, \mathbf{D}_{4h}, \mathbf{D}_{4d})$	$\mathbf{C}_{8v}(\mathbf{C}_8)$, $\mathbf{D}_8(\mathbf{C}_8)$, $\mathbf{D}_{4d}(\mathbf{S}_8)$, $\mathbf{D}_{8h}(\mathbf{C}_{8h})$
Decagonal	$\mathbf{C}_{10}$, $\mathbf{C}_{5h}$, $\mathbf{C}_{10v}$, $\mathbf{D}_{10}$, $\mathbf{C}_{10h}$, $\mathbf{D}_{5h}$, $\mathbf{D}_{10h}$	$\mathbf{C}_{10}(\mathbf{C}_5)$, $\mathbf{C}_{5h}(\mathbf{C}_5)$, $\mathbf{C}_{10v}(\mathbf{C}_{10}, \mathbf{C}_{5v})$, $\mathbf{D}_{10}(\mathbf{C}_{10}, \mathbf{D}_5)$, $\mathbf{C}_{10h}(\mathbf{C}_{10}, \mathbf{C}_{5h}, \mathbf{S}_{10})$, $\mathbf{D}_{5h}(\mathbf{D}_5, \mathbf{C}_{5v}, \mathbf{C}_{5h})$, $\mathbf{D}_{10h}(\mathbf{C}_{10h}, \mathbf{C}_{10v}, \mathbf{D}_{10}, \mathbf{D}_{5h}, \mathbf{D}_{5d})$	$\mathbf{C}_{10v}(\mathbf{C}_{10})$, $\mathbf{D}_{10}(\mathbf{C}_{10})$, $\mathbf{D}_{5h}(\mathbf{C}_{5h})$, $\mathbf{D}_{10h}(\mathbf{C}_{10h})$
Dodecagonal	$\mathbf{C}_{12}$, $\mathbf{S}_{12}$, $\mathbf{C}_{12v}$, $\mathbf{D}_{12}$, $\mathbf{C}_{12h}$, $\mathbf{D}_{6d}$, $\mathbf{D}_{12h}$	$\mathbf{C}_{12}(\mathbf{C}_6)$, $\mathbf{S}_{12}(\mathbf{C}_6)$, $\mathbf{C}_{12v}(\mathbf{C}_{12}, \mathbf{C}_{6v})$, $\mathbf{D}_{12}(\mathbf{C}_{12}, \mathbf{D}_6)$, $\mathbf{C}_{12h}(\mathbf{C}_{12}, \mathbf{C}_{6h}, \mathbf{S}_{12})$, $\mathbf{D}_{6d}(\mathbf{D}_6, \mathbf{C}_{6v}, \mathbf{S}_{12})$, $\mathbf{D}_{12h}(\mathbf{C}_{12h}, \mathbf{C}_{12v}, \mathbf{D}_{12}, \mathbf{D}_{6h}, \mathbf{D}_{6d})$	$\mathbf{C}_{12v}(\mathbf{C}_{12})$, $\mathbf{D}_{12}(\mathbf{C}_{12})$, $\mathbf{D}_{6d}(\mathbf{S}_{12})$, $\mathbf{D}_{12h}(\mathbf{C}_{12h})$
Icosahedral	$\mathbf{Y}$, $\mathbf{Y}_h$	$\mathbf{Y}_h(\mathbf{Y})$	–

magnetic moment vector in these classes is directed along the main axis.

In the case of the icosahedral symmetry, there is only one class $\mathbf{Y}_h$ ($\mathbf{Y}$) of type III. In this class, it is impossible (due to the presence of the operation *RI*) to find a direction for which the magnetic moment vector is invariant under all symmetry operations. However, due to different conditions (magnetostriction, external field, etc.), the icosahedral symmetry (the groups $\mathbf{Y}$ and $\mathbf{Y}_h$) can be reduced to the pentagonal groups $\mathbf{D}_5$ and $\mathbf{D}_{5d}$ if the action is along one of the fivefold axes or to the trigonal ($\mathbf{D}_3$ and $\mathbf{D}_{3d}$) and rhombic ($\mathbf{D}_2$ and $\mathbf{D}_{2h}$) groups if the action is along one of the three- or twofold axes. In this sense, the possibility of ferromagnetism in icosahedral quasicrystals is analogous to ferromagnetism in crystals with cubic symmetry. It is known that the lattice of the ferromagnetic phase of iron is not cubic (body-centered cubic), but tetragonal, with tetragonal distortion on the order of 10^{-5} [7, 8], which is too small to be observed experimentally. A distortion in the icosahedral quasicrystal (with the group $\mathbf{Y}_h$) due to magnetostriction along one of the five-order axes should reduce the symmetry to class $\mathbf{D}_{5d}$, which forms the ferromagnetic class $\mathbf{D}_{5h}$ ($\mathbf{S}_{10}$).

In quasicrystals, magnetostriction can generate phasons and, as a result, a sample becomes magnetically inhomogeneous. Actually, the experiments mentioned above have been explained in terms of large magnetic clusters with the size about 20 nm [2, 3]. Therefore, the magnetic state of these objects may be characterized as a “mictomagnetic” (mixed) one. For the mictomagnetic state, the susceptibility is analogous to an antiferromagnetic or spin glass state, but spontaneous magnetization after cooling in the field is typical of ferromagnets [8, 9]. In this sense, the “ferromagnetic” quasicrystals have many features in common with concentrated alloys CuMn and AuFe, where magnetic behavior can be described by the presence of large supermagnetic clusters with identical moments and anisotropy fields, but with random directions of the light magnetization axis. Upon increasing the concentration, creation of magnetic clusters becomes more probable and the long-range magnetic order can propagate over the entire sample [8, 9].

In conclusion, based on a magnetic-group analysis, we have predicted that ferromagnetism is incompatible with the icosahedral symmetry of quasicrystals. In magnetic field, the icosahedral symmetry is reduced to the pentagonal or trigonal or rhombic symmetry depending on the field direction. Magnetostriction can induce phason distortions in quasicrystals, and as a result, the system becomes magnetically inhomogeneous. Such a physical picture can explain the existing experimental data on “ferromagnetic” quasicrystals.

ACKNOWLEDGMENTS

Discussions with A. M. Polyakov were very helpful. We are very grateful to S. I. Simak for the discussions and help during the preparation of this work. Two of the authors (Yu. Kh. V. and E. 1.1.) thank the Russian Foundation for Basic Research (project no. 03-02-16970) and The Royal Swedish Academy of Sciences (KVA) for financial support.

REFERENCES

1. Z. M. Stadnik and G. Stroink, Phys. Rev. B **43**, 894 (1991).
2. *Physical Properties of Quasicrystals*, Ed. by Z. M. Stadnik (Springer, Berlin, 1999).
3. C. R. Lin, C. M. Lin, S. T. Lin, and I. S. Lyubutin, Phys. Lett. A **196**, 365 (1995).
4. I. S. Lyubutin, Ch. R. Lin, and S. T. Lin, Zh. Éksp. Teor. Fiz. **111**, 1449 (1997) [JETP **84**, 800 (1997)].
5. R. Lithshitz, Rev. Mod. Phys. **69**, 1181 (1997).
6. A. Yamamoto, Acta Crystallogr. A **52**, 509 (1996).
7. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media*, 2nd ed. (Nauka, Moscow, 1984; Pergamon, New York, 1984).
8. R. M. White, *Quantum Theory of Magnetism* (Springer, Berlin, 1983).
9. P. Gibbs, T. M. Harders, and L. H. Smith, J. Phys. F **15**, 213 (1985).

Journal of Experimental and Theoretical Physics, Vol. 100, No. 6, 2005, pp. 1129–1141.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 6, 2005, pp. 1282–1296.
Original English Text Copyright © 2005 by Berry, Smirnov.

ORDER, DISORDER, AND PHASE TRANSITIONS IN CONDENSED SYSTEMS

Phase Transitions in Simple Clusters¶

R. S. Berry[a] and B. M. Smirnov[b]

[a]*Department of Chemistry, University of Chicago 60637, Chicago, IL USA*
[b]*Institute for High Temperatures of Russian Academy of Sciences, Moscow, 127412 Russia*
e-mail: smirnov@oivtran.iitp.ru

Received January 10, 2005

Abstract—Formation of the liquid state of clusters with pairwise interactions between atoms is examined within the framework of the void model, in which configurational excitation of atoms results from formation of voids. Void parameters are found from computer simulation by molecular dynamics methods for Lennard-Jones clusters. From that standpoint, phase transitions are analyzed in terms of two aggregate states. This information allows us to divide the entropy jump during a solid-liquid phase transition into two parts: one corresponds to configurational excitation at zero temperature and the other arises from thermal vibrations of atoms. The latter part contributes approximately 40% for Lennard-Jones clusters consisting of 13 and 55 atoms, increasing to 56% for bulk inert gases. These magnitudes explain the validity of melting criteria based on thermal motion of atoms, even though the distinctive mechanism of this phase transition results from configurational excitations. It is shown that the void concept allows analyzing various aspects of the liquid state of clusters including the existence of a limiting freezing temperature below which no metastable liquid state exists, as well as the existence and properties of glassy states that may exist below the freezing limit. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Clusters, being systems of relatively small finite numbers of bound atoms, differ from macroscopic atomic systems in several properties. Their solid states are characterized by sharp, nonmonotonic dependence of their population on the number of component atoms. The most striking aspect is the occurrence of “magic numbers” that corresponds to completed atomic shells, often of polyhedra rather than lattices. In experiments, these favorable structures exhibit heightened populations and stabilities, see, e.g., [1–3]. In the subsequent analysis, we focus on clusters with pairwise interactions between atoms. In this case, the pairwise character and magnitude of the interaction means that the interaction energies between atoms of a cluster are small compared with a typical electronic excitation energy. This criterion is valid for clusters of inert gas atoms and clusters of molecules typically found in a gaseous state under normal conditions. (It does not hold for covalently bound clusters or metallic clusters.) We consider the phase change of clusters between the solid and liquid states and examine how it differs in principle from the traditional melting/freezing transition in macroscopic systems. Indeed, the phase transition in macroscopic systems proceeds by a sharp step in specific properties, and hence, in classical thermodynamics, there are phase transitions of the first and second orders [4–9], depending on the behavior of the derivatives of specific thermodynamical quantities. In the cluster case, computer simulation exhibits coexistence of the solid and liquid phases [10–13] over some finite temperature and pressure band that makes the phase change of clusters richer than for macroscopic systems, and, in a sense, makes the classical thermodynamic classification of phase transitions, based on specific discontinuities, inapplicable here, even while the basic laws of thermodynamics remain completely valid. Because of the coexistence of aggregate states in the phase change of clusters, dividing the phase transitions into sharply divided types loses its sense.

The phase transition is a collective phenomenon that results from simultaneous interaction of many atoms. Therefore, simple analytical one-particle models are not suitable for its analysis. In considering the phase change in clusters, we use the results of computer simulation for clusters whose atoms interact through the Lennard-Jones potential. We focus mainly on the liquid state of these systems, whose properties depend monotonically on the number of cluster atoms, in contrast to the solid state, which exhibits its irregular dependence (with magic numbers) on the number of cluster atoms. As a result of melting, the crystalline distribution of atoms is lost, and the liquid state has more or less amorphous structure, although a shell-like distribution of atoms may be conserved to some degree. The aim of this paper is the analysis of some properties of liquid cluster systems on the basis of appropriate models. In constructing our model, we must take the specifics of the cluster liquid state into account. In general, this can follow from the probability of the total kinetic energy of cluster atoms held at constant energy, for example, as was done in [11] for the Lennard-Jones cluster of

---

¶ This article was submitted by the authors in English.

1063-7761/05/10006-1129$26.00 © 2005 Pleiades Publishing, Inc.

13 atoms. Here, this probability has a bimodal form in a range of internal energies near the state of classical melting, i.e., where the chemical potentials of the solid and liquid states are equal. (For convenience, we here refer to this state as the "melting point.") The two maxima of this distribution and their vicinities correspond to two aggregate states, solid and liquid.

In modeling a large liquid cluster with pair interactions between atoms by a spherical liquid drop, we take into account that atoms in this aggregate state occupy a larger volume than in the solid state. Then we can consider the cluster's transition from solid to liquid as a result of formation of voids inside the cluster [14]. We consider a void as an elementary configuration excitation. A void is a perturbed, even a relaxed, vacancy; in contrast to a vacancy in a solid, a void has an indefinite volume and shape that changes in time. On the basis of computer simulation results, we find the average void parameters as they emerge for macroscopic inert gases [15–19] on the basis of their measured parameters. This allows us to analyze various aspects of the phase transitions in condensed inert gases as well as in Lennard-Jones clusters.

Introduction of a void as an elementary configuration excitation is in reality a simplification of a general analysis of the potential energy surface of an ensemble of interacting atoms [20]. In a multidimensional space of atomic coordinates, the potential energy surface for an atomic ensemble with pairwise interactions consists of many potential wells separated by saddles [13, 21–24]. Evolution of this ensemble is described by saddle-crossing dynamics [23]. In particular, a cluster is found near a minimum of the potential energy long enough to equilibrate its vibrations; transition to a neighboring minimum typically proceeds relatively slowly by comparison [25]. (Of course, if there are minima separated by low barriers, equilibration among these may occur on the same time scale as vibrational relaxation, and can be treated appropriately.) First, this leads to a short-range, short-duration order for any amorphous structure of atoms, because each configuration of atoms is preserved for a relatively long time (such that the atomic kinetic energy is not very large). Second, this allows us to introduce an average void by averaging parameters over times that exceed the typical vibrational period but are shorter than typical well-to-well passages. This allows us to separate configuration excitation that is responsible for the phase transition from vibrational excitation associated with an increase of the kinetic energy of the atoms. This fact is of importance for the analysis of the phase transition.

We note that the phase change in clusters is richer and more complicated than in bulk systems. In particular, the sharp onset of a liquid state is absent for Lennard-Jones clusters of 8 and 14 atoms (and others) [26]; these are examples of systems that do not show bimodal distributions of kinetic energies on time scales of vibrational relaxation. In addition, large clusters can exhibit several aggregate states associated with melting of various atomic shells [27, 28]. Below, we focus on clusters with completed atomic shells and consider melting of surface shells. In these cases, the melting process is clearer and more easily distinguished. First, because the surface shell contains a considerable fraction of the cluster's atoms, the statistical weight of configurationally excited states with voids in the surface shell is relatively large, as is the entropy of transition to this state. This leads to a stable liquid state for the surface layer of these clusters, as well as a state composed entirely of liquid. Next, in contrast to the liquid state of bulk systems, where a void differs clearly from the vacancy in a solid, a surface void in a not-so-large cluster can be considered a perturbed vacancy. For example, the energy of formation of a void in the liquid state of a bulk inert gas is approximately half the energy of formation of a vacancy in the solid state of a bulk inert gas [15–17]. In the case of clusters under consideration, the energy difference of formation of surface vacancies and voids is not very large, and hence a void can be considered a perturbed vacancy. This facilitates the analysis.

Because some concepts of classical thermodynamics of macroscopic systems are not valid for clusters, it is necessary to revise some of those concepts in order to apply them to clusters. In particular, the phase of an aggregate state is defined in classical thermodynamics as a uniform distribution of atoms that is restricted by boundaries. From the standpoint of the void concept, this means that the liquid aggregate state as a configurationally excited state has to contain many elementary excitations—voids. To transfer this concept to small clusters, it is necessary to revise the definition, to be done in what follows.

The void concept gives a sufficiently detailed picture of the liquid aggregate state, in comparison with classical thermodynamics, and allows us to resolve its apparent paradoxes. As an example, we consider the criterion of the phase transition for macroscopic systems. According to the widely used Lindemann criterion [29, 30], bulk melting proceeds at the temperature at which the ratio of the atomic oscillation amplitude to the distance between nearest neighbors reaches a certain value (10–15%). Development of numerical methods for computer cluster simulation gave new variants of this criterion of cluster melting on the basis of the Etters–Kaelberer parameter [31–33] or Berry parameter [12, 34]; these account for pair correlations in positions of atoms. These parameters have jumps at melting and, as with the Lindemann criterion, are connected with thermal motion of atoms, even though the melting results from configurational excitation. Hence, there is an apparent contradiction between the nature of the phase transition in ensembles of bound atoms due to configurational excitation and the practical criteria for this transition based on thermal motion of atoms. The subsequent analysis of numerical parameters of this phase transition exhibits the resolution of this apparent contradiction.

Although vibrational excitation of finite and "infinite" numbers of bound atoms that characterizes thermal motion of atoms is separated from configuration excitation that is responsible for the phase transition, thermal motion of atoms gives a contribution to the entropy jump ΔS at the melting point, because the solid state is more compact than the liquid aggregate state. Reflecting just the thermal (kinetic energy) part of the entropy jump, the vibrational parameters nevertheless simultaneously characterize the phase transition. By exhibiting the connection between the configurational and kinetic-energy aspects, the void analysis justifies applying the melting criteria based on thermal motion of atoms.

The goal of this paper is the analysis of Lennard - Jones clusters with completed shells from the standpoint of the void concept. This consists in obtaining the void parameters from treatment of numerical computer calculations and in the analysis of these data to connect the two kinds of excitation.

2. CHARACTER OF CLUSTER CONFIGURATIONAL EXCITATION

We consider an ensemble of interacting atoms in which the ground electronic state is well separated from electronically excited states, and these excited states do not partake in evolution of the atomic ensemble, i.e., development of this system can be described in terms of motion on the potential energy surface (PES) in a multidimensional space of nuclear coordinates, and this PES corresponds to the electron ground state. The PES has many local minima, which was discovered in first numerical calculations of the cluster energy at zero temperature for a simple character of atomic interactions [35–37]. Early algorithms that made oversimplifying assumptions about the potential landscape led to underestimations of the number of minima, but with increasing the computing power, more elaborate methods made it possible to explore these landscapes fairly thoroughly and to obtain plausible estimates regarding the dependence of the number of minima on the number n of atoms in the cluster. For Lennard-Jones clusters, the number of geometrically distinct minima increases somewhat faster than exponentially with n; there are roughly $n!$ permutational isomers of each of these, and hence the total number increases roughly as $n!\exp(an)$ [21, 36–39]. As a result, cluster evolution consists of transitions between neighboring local minima of PES that correspond to saddle-crossing dynamics [13, 22–24]. This concept is a basis for investigation of various ensembles of interacting atoms, from simple clusters to biological molecules [20, 24].

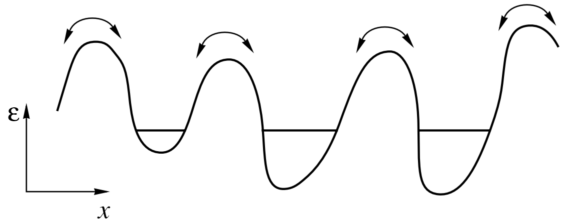


Fig. 1. The evolution of an ensemble of atoms in the ground electron state as the propagation of a point in the phase space of atomic coordinates resulting from transition between neighboring local minima of the potential energy surface.

The character of transitions between neighboring minima of PES is given in Fig. 1, which shows schematic projections of a PES on planes in the space of atomic coordinates where only the coordinate related to a specific transition between two local minima of the PES varies. (These planes are different for each transition.) Energy levels for each well indicate an average atomic energy along the coordinate of the transition. Because this energy is significantly less than the barrier height, such transitions proceed seldom, only when the kinetic energy of atoms in the transition degree of freedom exceeds its average energy adequately. Hence, the system has many oscillations inside a given well until it transfers to another local minimum of the PES. Then identifying a given local (but not global) minimum of the PES as a configurational excitation of the system of interacting atoms, we can separate it from thermal motion associated with atomic oscillations. This allows us to consider the configurational excitation independently of vibrational excitations and is the basis of the following analysis. Moreover, taking a realistic assumption that a typical time of establishment of the thermodynamic equilibrium for thermal motion of atoms is short compared to a typical time of transition between local minima of the PES, we can introduce the temperature of thermal motion of bound atoms for each configuration excitation.

We use an approach based on the assumption of a large number of local minima on the PES. In particular, the Lennard- Jones cluster of 13 atoms is characterized by 1478 local minima and 17357 saddle points of the potential energy surface [22]; early estimates found only 988 local minima [35, 36]. As noted above, the number of geometrically distinct local minima increases roughly exponentially with increasing the number n of atoms in the cluster [20, 21, 38]. We construct an aggregate state from configurationally excited states. Restricting ourselves to the local minima of the PES that are occupied with a nonnegligible probability in a range of cluster temperatures under consideration, we join the local minima with similar excitation energies into one aggregate state. Because of the large number of such local minima, the aggregate state is characterized by a large statistical weight and, correspondingly, by a large entropy, and hence the probability for a cluster to be found in this aggregate state may be considerable, even when the excitation energy is signifi-

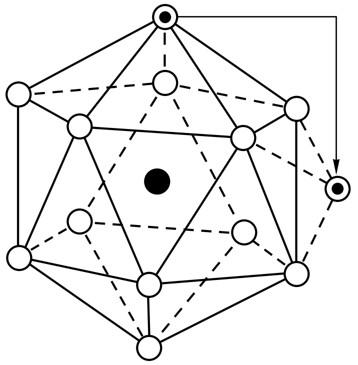


Fig. 2. The character of the lowest configurational excitation in the icosahedral cluster consisting of 13 atoms at zero temperature.

cantly less than *kT*. Below, we illustrate this definition of the cluster aggregate state with examples.

3. CONFIGURATIONAL EXCITATION OF THE ICOSAHEDRAL CLUSTER OF 13 ATOMS

Dividing cluster excitations into two parts, configurational and vibrational, we consider these parts independent. Next, for configurational excitation, we use the void model, considering this excitation as a result of formation of voids. Then we can express the parameters of the phase transition and other cluster properties through the parameters of forming voids. We restrict ourselves to clusters with completed shells, which simplifies this analysis because a void, an elementary configurational excitation, is in reality a perturbed vacancy.

We start the analysis of configurational excitation from the simplest cluster with completed shells that has the icosahedral structure and consists of 13 atoms. In the lowest-energy state, its first (and only) shell is filled. Configurational excitation of this cluster consists in transition of one atom from the surface shell to the cluster surface as shown in Fig. 2. After formation of a vacancy on the cluster shell, the atoms around the vacancy are distributed over a larger space due to thermal motion, and the promoted atom moves over the cluster surface more freely than any of the other atoms. As a result of the configurational transition, an excited cluster state has the statistical weight g and the additional entropy ΔS_0 given by

$$g = 12 \times 15 = 180, \quad \Delta S_0 = \ln g = 5.2. \qquad (1)$$

Here, the value 12 is the number of shell atoms, any of which can be promoted, and 15 is the number of positions for a promoted atom if it is not in a site bordering the new vacancy. It is important that thermal motion of atoms gives a contribution to these values near the melting point because of the free motion of bound atoms in this configurationally excited state.

We now determine energetic parameters of this transition. In the ground state, each surface atom of this icosahedral cluster has five atoms from the surface shell and the central atom as nearest neighbors. The pairwise interactions therefore yield six “bonds” to these atoms. When this atom is located on the cluster surface, it has only three nearest neighbors. Then the energy of this atomic transition onto the hollow between three atoms on the cluster surface is roughly equal to 3*D*, where *D* is the energy needed to break one bond. (We assume atoms to be classical, and hence the depth of the potential well *D* for the two-atom interaction coincides with the dissociation energy of a diatomic molecule.) Next, if a surface atom is transferred to a neighboring position on the cluster surface, it retains two bonds between nearest neighbors as it makes the transition between neighboring hollows on the cluster surface. Hence, for this transition, an atom must overcome a barrier whose magnitude is roughly 1*D*. Figure 3 gives the energies of these states at zero temperature and the values of barriers that separate them for the Lennard-Jones interaction potential between atoms [40]. We see that the difference between the data in Fig. 3 and the above values is not substantial. Hence, our consideration of an elementary configurational excitation, a void, as a perturbed vacancy is justified.

Figure 3 gives the energies of these states at zero temperature and the values of barriers that separate them [40]. As a result of configurational excitation, an atom transfers from the shell of 12 atoms, as shown in Fig. 2. For this transition, an atom must overcome a barrier; likewise, transitions to other positions on the cluster surface are accompanied by overcoming energetic barriers. Increasing the energy facilitates transitions between different stable positions on the cluster surface, as well as exchanges between a configurationally excited atom and another one. All configurationally excited states with promotion of one atom are connected in the liquid state, and hence the system may find all permutations among the atoms of any attainable structure. This follows from the data in Fig. 3. Thus, all the configurational states with one atom promoted from the outer shell comprise the lowest-energy excited aggregate state.

We now analyze the character of configurational excitation of this cluster in the phase coexistence range where thermal motion of atoms influences the transition parameters. We base this on the results of computer simulation of the Lennard-Jones cluster of 13 atoms [11], in which this cluster is considered a member of a microcanonical ensemble [41], i.e., the total cluster energy is conserved during the cluster’s evolution. In a particular band of energies, the probability distribution of the total kinetic energy (or mean kinetic energy per atom) of the cluster becomes the bimodal distribution, which confirms the existence of the aggregate states in the dynamic equilibrium in this case (solid and liquid). Therefore, in this range, we can treat the results of computer simulation [11] in terms of a dynamic equilibrium

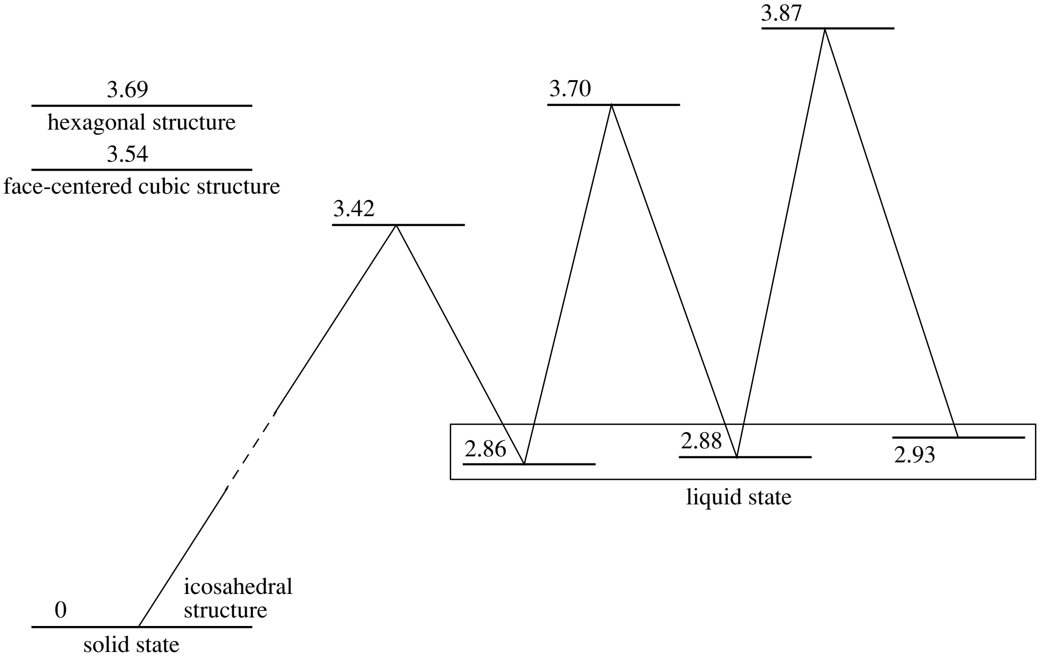


Fig. 3. The energy and barrier positions for the lowest configurational excitations of the Lennard-Jones cluster consisting of 13 atoms according to [40].

of two aggregate states. This was fulfilled partially in [42–44], and we give the results of this treatment below.

We now give general formulas of this consideration in which the results of numerical calculations are compared with simple formulas that reflect a simple concept. For the classical motion of bound atoms inside the cluster, we represent the energy E of a cluster consisting of n atoms, with a pair interaction between them, in the form

$$E = U + K = \sum_{i,j} u(\mathbf{r}_{ij}) + \frac{m}{2}\sum_i \left(\frac{d\mathbf{r}_i}{dt}\right)^2. \tag{2}$$

Here, U is the total potential energy, K is the total kinetic energy of atoms, $u(\mathbf{r}_{ij})$ is the pair interaction potential between atoms at a distance

$$\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j,$$

where $\mathbf{r}_i$ and $\mathbf{r}_j$, are the atomic coordinates, and m is the atomic mass. This formula is the basis of our analysis of cluster computer simulations. We consider the properties of two terms of this formula, taking into account that thermal equilibrium is usually established for atomic vibrations. This allows us to introduce the atomic temperature T, with the motion of atoms treated as that of a set of harmonic oscillators. The cluster temperature is defined from the relation

$$K = \frac{3}{2}nT, \tag{3}$$

where $n \gg 1$ and the total kinetic energy of atoms is averaged over times much longer than the period of atomic oscillation. A typical oscillation time τ for atoms can be expressed through the Debye frequency ω_D as

$$\tau \sim \frac{1}{\omega_D} \sim \frac{\hbar}{D}. \tag{4}$$

We note that for a microcanonical ensemble, this definition of the effective temperature, although useful and widely chosen, is not the only one, and different definitions are not equivalent [45].

The global minimum of the potential energy surface of this cluster has the structure of a regular icosahedron, and it corresponds to the cluster's ground state, its equilibrium state at zero temperature. Transitions from the global minimum to other local minima of the potential energy surface for a 13-atom cluster are responsible for the phase transition to the liquid aggregate state. Figure 3 gives the energies of the lowest configurationally excited states of this cluster at zero temperature and the values of barriers that separate them [40]. As a result of configurational excitation, an atom transfers from the shell of 12 atoms, as shown in Fig. 2. At high temperatures, the lowest configurationally excited state is the liquid aggregate state, which is to be justified below.

The energy of an isolated cluster of 13 atoms in the energy range where both solid and liquid may be stable can be represent as

$$E = -E_0 + E_{\mathrm{ex}} = U^k_{\mathrm{sol}} + K_{\mathrm{sol}} = \Delta E + U^k_{\mathrm{liq}} + K_{\mathrm{liq}}, \tag{5}$$

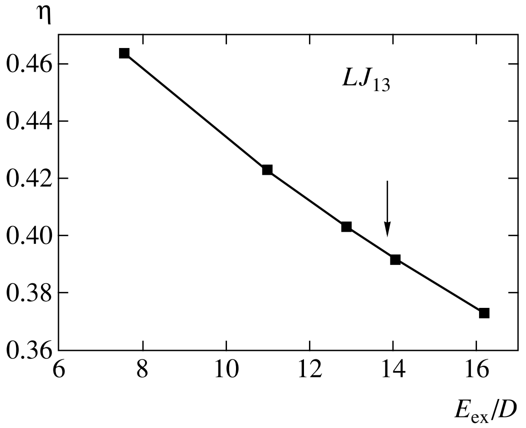


Fig. 4. The dependence on the excitation energy for the energy part related to the kinetic energy of atoms for an isolated Lennard-Jones cluster of 13 atoms. This value is identical for the solid and liquid cluster states. The arrow indicates the excitation energy of the phase transition $w_{sol} = w_{liq}$.

where E_0 is the binding energy of cluster atoms at zero temperature, E_{ex} is the excitation energy, K_{sol} and K_{liq} are the total kinetic energies of atoms for the solid and liquid cluster states, U^k_{sol} and U^k_{liq} are the average potential energies of the cluster for the kth local minimum of the potential energy surface, and ΔE is the average excitation energy above the energy needed to reach U^k_{liq} required to produce the labile liquid aggregate state. The values in this formula are averaged over times that exceed a typical time of atom oscillations but are less than a typical transition time between cluster aggregate states. Within the framework of this consideration, we join energetically similar local minima of the cluster potential energy into one aggregate state, assuming the transitions between local minima of the same aggregate state to be more effective than transitions between states that belong to different aggregate states. In other words, we assume that the system, although liquid, explores the local minima available to it far more frequently than it returns to the solid with which it is in the dynamic equilibrium-on long enough time scales.

We introduce the effective temperature for a given aggregate state of the cluster on the basis of a formula transformed from formula (3):

$$T = \frac{2}{3n-6}K = \frac{2K}{33}, \qquad (6)$$

where $n = 13$ is the number of cluster atoms, and the energy fraction η related to the kinetic energy of atoms is

$$\eta_{sol} = \frac{K_{sol}}{E_{ex}}, \quad \eta_{liq} = \frac{K_{liq}}{E_{ex} - \Delta E}. \qquad (7)$$

If atomic motion is a combination of harmonic oscillators, we have $\eta = 0.5$. Anharmonicity of the oscillations typically leads to a decrease of this value, and $\eta(E_{ex})$ decreases with an increase of E_{ex}. Treatment of the results of computer simulation [11] for this cluster by the method in [43] gives the dependence $\eta(E_{ex})$ that is represented in Fig. 4. We note that

$$\eta_{sol}(E_{ex}) = \eta_{liq}(E_{ex})$$

within the limits of the result accuracy, while this quantity has different values for the solid and liquid states at identical temperatures. This value starts from

$$\eta(E_{ex} = 0) = 0.5,$$

where the system can be described in terms of harmonic oscillators and decreases monotonically with increasing the excitation energy because of the increasing role of anharmonicity of the dominant stretching modes.[1] Hence, the parameter η characterizes the influence of the anharmonicity in atomic motion of an isolated cluster as the excitation energy increases.

From these data, we have the excitation energy of the cluster liquid state,

$$\begin{aligned} \Delta E &= \frac{K_{sol} - K_{liq}}{\eta(E_{ex})} \\ &= E_{ex}\left(1 - \frac{K_{liq}}{K_{sol}}\right) = 2.49 \pm 0.05. \end{aligned} \qquad (8)$$

Comparison of formula (8) with the data in Fig. 3 for the excitation energies at zero energy shows that the difference of these energies is not significant. In contrast to this, the ratio of the energy of void formation for bulk inert gases to the vacancy energy formation is approximately one half. The energy of void formation at the melting point $T_m = 0.29D$ allows us to find the entropy jump of the bulk system at the melting point

$$\Delta S_m = \frac{\Delta E}{T_m} = 8.6 \pm 0.2. \qquad (9)$$

Comparing this with the entropy jump ΔS_0 of the cluster at zero temperature according to formula (1), we find

$$\frac{\Delta S_0}{\Delta S_m} \approx 0.6. \qquad (10)$$

---

[1] Some modes, notably bending modes in molecules and, at high energies, high-frequency phonons in large clusters and solids, may have negative anharmonicities. For such systems, this line of reasoning requires further scrutiny. But for clusters of at least several hundred atoms, it is a valid assumption that the anharmonicity reduces the spacing of the relevant modes as the energy increases.

Thus, the different character of atom motion in the solid compact aggregate state and in the liquid aggregate state with its sparser distribution of atoms increases the entropy jump.

An isolated cluster, viewed as a microcanonical ensemble of atoms, is characterized by two temperatures if it can be found in two aggregate states only. These temperatures are determined by formula (6) for each aggregate state, and ignoring the anharmonicity, which gives $\eta = 0.5$, we use formula (8) to obtain the difference of the atomic temperatures $T_{\rm sol}$ and $T_{\rm liq}$ of the solid and liquid aggregate states near the melting point:

$$T_{\rm sol} - T_{\rm liq} = \frac{2\Delta E}{33} \approx 0.15D. \qquad (11)$$

Figure 5 gives the values of these temperatures for the Lennard-Jones cluster of 13 atoms as a function of the cluster excitation energy, and these data follow from treatment of the computer simulation results [11] for this cluster. Along with these temperatures, we can introduce the configurational cluster temperature $T_{\rm ef}$ that follows from the equilibrium between the solid and liquid cluster states according to the formula

$$\frac{w_{\rm liq}}{w_{\rm sol}} = \exp\left(-\frac{\Delta F}{T_{\rm ef}}\right) = \exp\left(-\frac{\Delta E}{T_{\rm ef}} + \Delta S\right), \qquad (12)$$

where $w_{\rm sol}$ and $w_{\rm liq}$ are the respective probabilities for the cluster to be found in the solid or liquid states and ΔF is the free energy jump at melting. The configurational temperature is determined by populations of the solid and liquid aggregate states. Figure 5 gives the dependence of the configurational temperature on the cluster excitation energy. The configurational temperature tends to the solid temperature in the limit of low temperatures, and to the liquid temperature in the limit of high temperatures.

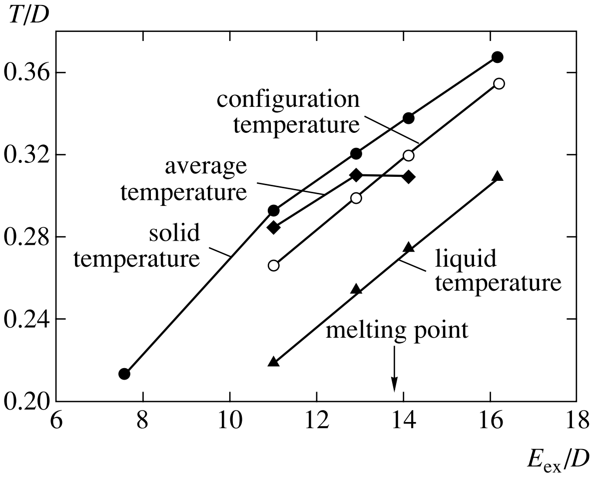


Fig. 5. Temperatures of the Lennard-Jones cluster of 13 atoms depending on the excitation energy (the temperatures of the solid and liquid aggregate states, average transversal temperature, and configuration temperature).

4. CONFIGURATIONAL EXCITATION OF THE ICOSAHEDRAL CLUSTER OF 55 ATOMS

We consider a cluster as a member of a canonical ensemble [41], when it is maintained under isothermal conditions. Experimentally [46–49], this condition can be reached when clusters are located in a gas of light (usually, helium) atoms that collide with clusters and metallic walls maintained at a certain temperature. Collisions establish the wall temperature for each cluster if the typical time for equilibration to a gas temperature as a result of collisions with the atomic heat bath is short compared to a typical dwell time of the cluster in each aggregate state. As earlier, we assume the existence of the two aggregate states [42–44] and use the dynamic coexistence of phases in clusters [11, 50–52] within a temperature range; in other words, we assume that part of the time, the cluster is found in one aggregate state and in the remainder, it is found in the other. In addition, while the cluster is in each aggregate state, vibrational equilibrium is established [25], and hence the temperature of bound atoms for each aggregate state coincides with the thermostat temperature if the cluster is in a canonical ensemble [41]. In this case, the probability $w_{\rm liq}$ that the cluster is found in the liquid state is given by the formula [18, 43, 44]

$$\begin{aligned} w_{\rm liq} &= \frac{p}{1+p}, \qquad (13) \\ p &= \exp\left(-\frac{\Delta E}{T}\right) = \exp\left(\Delta S - \frac{\Delta E}{T}\right), \end{aligned}$$

where T is the cluster temperature (which coincides with the thermostat temperature), ΔE is the energy of configurational excitation, ΔS is the entropy jump as a result of melting, and ΔF is a change of the free energy. The parameters of the phase transition ΔE and ΔS determine the behavior of the cluster heat capacity, which we connect with the cluster heat capacity that can be calculated from computer simulations. Using the average kinetic energy of atoms for each aggregate state, we characterize each of those states by the mean potential energy, i.e., we ignore the broadening of the energy of each cluster state due to fluctuations.

We first determine the cluster heat capacity and separate its "resonance" part, the peak due to the phase transition. According to formula (5), we have

$$E = -E_0 + \frac{K_{\rm sol}}{\eta_{\rm sol}} w_{\rm sol} + \Delta E w_{\rm liq} + \frac{K_{\rm liq}}{\eta_{\rm liq}} w_{\rm liq}, \qquad (14)$$

and because the cluster is in a thermostat, $K_{\rm sol} = K_{\rm liq}$. Assuming that $\eta_{\rm sol} = \eta_{\rm liq}$, we obtain the average cluster energy according to formula (14),

$$E = \frac{K_{\text{sol}}}{\eta_{\text{sol}}} + \Delta E w_{\text{liq}} = \frac{K_{\text{sol}}}{\eta_{\text{sol}}} + \Delta E \frac{p}{1+p}. \tag{15}$$

We first consider the Lennard-Jones cluster of 13 atoms in a thermostat. We note that the anharmonicity of atomic oscillations affects the degree of the configurational excitation. Indeed, because the anharmonicity is greater for the liquid than for the solid cluster, the density of vibrational states increases faster with energy than does that of the solid, and therefore the average potential energy for the isothermal liquid cluster is higher than that for the solid. If the melting temperatures for the isolated and isothermal clusters are approximately equal (as they are for clusters of ca. 100 atoms or more), we find a special excess change of the cluster's potential energy in the isothermal case that does not appear in the constant-energy case. Taking the melting point $T_{\text{m}} = 0.29D$ for both cases and the corresponding kinetic energies at this temperature to be

$$K_{\text{sol}}(T_{\text{m}}) = K_{\text{liq}}(T_{\text{m}}) \approx 2.9D,$$

we find the excess excitation energy $\Delta E'$ as a change of the average potential energy per atom given by

$$\begin{gathered} \Delta E' = \Delta E + \frac{K_{\text{liq}}(T_{\text{m}})}{\eta_{\text{liq}}(T_{\text{m}})} - \frac{K_{\text{sol}}(T_{\text{m}})}{\eta_{\text{sol}}(T_{\text{m}})} = \Delta E \\ + K_{\text{sol}}(T_{\text{m}})\left[\frac{1}{\eta(15.3D)} - \frac{1}{\eta(11.7D)}\right] = 3.1D, \end{gathered} \tag{16}$$

where $\Delta E = 2.5D$ is the energy difference for aggregate states of an isolated cluster at the melting point. We here assume that the anharmonicity parameter η is identical for both aggregate states of an isolated cluster, and its dependence on the excitation energy is given in Fig. 4. Next, the dependence of the temperature of a given aggregate state on the excitation energy is represented in Fig. 5, and, as indicated in formula (16), the excitation energies of isolated clusters correspond to the temperature $0.29D$ for each aggregate state. Therefore, the anharmonicity of the solid aggregate state under isothermal conditions is higher than that for an isolated cluster, whereas for the liquid state, we have a different relation between these values. As a result, the isothermal phase transition requires a greater change of potential energy than that at constant energy, due to interactions between atoms for the Lennard-Jones 13-atom cluster. Due to the anharmonicity, the energy change for an isothermal cluster exceeds that of the isolated cluster approximately by 20%.

We next analyze the temperature dependence of the heat capacity under isothermal conditions. We represent the cluster heat capacity in the isothermal case as [53, 54]

$$C = \frac{dE}{dT} = C_0 + \frac{d(\Delta E w_{\text{liq}})}{dT}, \tag{17}$$

where the first term characterizes the cluster's heat capacity in the absence of the phase transition,

$$C_0 = \frac{d(K_{\text{sol}}/\eta_{\text{sol}})}{dT} = \frac{d(K_{\text{liq}}/\eta_{\text{liq}})}{dT}, \tag{18}$$

and the resonance part of the heat capacity is

$$C_{\text{res}} = \frac{d(\Delta E w_{\text{liq}})}{dT} = \frac{\Delta E^2}{T^2}\frac{p}{(1+p)^2}. \tag{19}$$

We here assume that the energy of configurational excitation ΔE and the entropy jump ΔS are independent of the temperature. Formula (19) leads to the maximum $C_{\text{res}}^{\max}$ at the melting point T_{m} defined in this case as $p(T_{\text{m}}) = 1$. We have

$$C_{\text{res}}^{\max} = \frac{\Delta E^2}{4T_{\text{m}}^2} = \frac{\Delta S^2}{4}. \tag{20}$$

To account for the temperature dependence of the entropy jump, we represent it as

$$\Delta S = \Delta S_0 + aT, \tag{21}$$

where ΔS_0 is the entropy jump at zero temperature. At the melting point, this gives

$$a = \frac{\Delta S_{\text{m}} - \Delta S_0}{T_{\text{m}}}, \tag{22}$$

where ΔS_{m} is the entropy jump at the melting point. On the basis of formulas (13) and (19), we obtain the resonant part of the heat capacity:

$$\begin{gathered} C_{\text{res}}^{\max} = \frac{\Delta E}{(1+p)^2}\frac{dp}{dT} = \Delta E\frac{p}{(1+p)^2} \\ \times\left(\frac{d\Delta S}{dT} + \frac{\Delta E}{T_{\text{m}}^2}\right) = \frac{\Delta S_m^2}{2} - \frac{1}{4}\Delta S_0 \Delta S_{\text{m}}. \end{gathered} \tag{23}$$

Because numerical calculations by the molecular dynamics method allow one to determine the heat capacity maximum, this relation can be used for evaluating the entropy at zero temperature as

$$\Delta S_0 = 2\Delta S_{\text{m}} - \frac{4C_{\text{res}}^{\max}}{\Delta S_{\text{m}}}. \tag{24}$$

We now use this formula for the Lennard-Jones cluster of 55 atoms. As our basis, we take computer simulations of this cluster in [28, 55, 56], which give the parameters characterizing the phase transition within the ranges

$$\Delta E = 15 \pm 1, \quad \frac{T_{\mathrm{m}}}{D} = 0.31 \pm 0.01,$$

$$C_{\mathrm{res}}^{\max} = 650 \pm 50.$$

From this, we have

$$\Delta S_{\mathrm{m}} = \frac{\Delta E}{T_{\mathrm{m}}} = 48 \pm 5. \tag{25}$$

On the basis of formula (24), we have

$$\Delta S_0 = 36 \pm 15. \tag{26}$$

The large uncertainty here makes this result relatively uninformative. We need a way to do better.

We now determine the entropy jump at zero temperature from another standpoint. The energy of formation of one vacancy at zero temperature can be found by comparing the total binding energies of atoms ε_{55} and ε_{56} for the Lennard-Jones clusters of 55 and 56 atoms. On the basis of calculations in [57] for ε_{55} and ε_{56}, we have [43]

$$\Delta\varepsilon = \varepsilon_{56} - \varepsilon_{55} = 2.64D$$

at zero temperature. The direct calculations for lower excitations of this cluster [58] lead to the minimal excitation energy 2.63D. The proximity of these values confirms that an excited atom transferred onto the cluster surface can be treated as being well removed from the vacancy from which it came. Evidently, the energy of void formation, the relaxed form, is less than the energy $\Delta\varepsilon$ of formation of the initial vacancy. This implies that the number of atoms leaving the body of the cluster to form voids is

$$v \geq \frac{\Delta E}{\Delta\varepsilon} \approx 5. \tag{27}$$

Taking the number of voids in the liquid state of the Lennard-Jones cluster of 55 atoms to be $v = 5$–7, we determine the entropy jump at zero temperature ΔS_0 as we treat the solid-liquid transition to be a consequence of transitions of atoms from the outer cluster shell onto its surface. Because of the icosahedral structure of this cluster, its outermost shell consists of 42 atoms, and there are 80 positions with 3-atom “hollows” on the surface for atoms promoted from the outer shell. A new vacancy on the cluster edge or surface has $l = 6$ neighboring atoms, and a vertex vacancy has only $l = 5$. Therefore, if v atoms transfer onto the cluster surface, then vl bonds are lost in the cluster surface for atoms transferred to any of the 80 positions on the cluster surface if we assume that v transferring atoms on the cluster surface do not border vacancies on the cluster shell. From this, for the configurational excitation of the cluster at zero temperature, we find the entropy jump that results from v atoms moving from the outermost shell,

$$\Delta S_0 = \ln C_m^v C_{42}^v,$$

where $m = 80 - vl$ is the number of positions on the cluster surface for transition of atoms from the outermost cluster shell. This formula implies that the entropy jump at zero temperature is

$$\Delta S_0 = 28.5 \pm 0.3 \quad \text{for} \quad v = 5,$$

$$\Delta S_0 = 31.6 \pm 0.4 \quad \text{for} \quad v = 6,$$

$$\Delta S_0 = 32.3 \pm 0.7 \quad \text{for} \quad v = 7.$$

Thus, the entropy jump at zero temperature depends weakly on the number of transferred atoms, and the average value of the entropy jump at zero temperature is

$$\Delta S_0 = 31 \pm 2. \tag{28}$$

One can see that this value is well within the range in formula (26). Because the accuracy is higher in this case than in formula (26), we use formula (24) for determination of ΔS_{m}. Then formula (24) gives

$$\Delta S_m = \frac{\Delta S_0}{4} + \sqrt{\frac{\Delta S_0^2}{16} + 2C_{\mathrm{res}}^{\max}}, \tag{29}$$

and on the basis of formula (28) and the calculated maximal heat capacity

$$C_{\mathrm{res}}^{\max} = 650 \pm 50,$$

we obtain

$$\Delta S_m = \frac{\Delta E}{T_m} = 45 \pm 2. \tag{30}$$

This result, together with its validity range, is consistent with formula (25), but we now have a result with greater precision and presumably with greater accuracy. Thus the analysis of computer simulations of these clusters by molecular dynamics allows us to determine some thermodynamical parameters of the phase transition within the framework of a simple scheme. Below, we analyze these results together with the microscopic nature of the phase transitions.

5. CHARACTER OF PHASE TRANSITIONS IN SIMPLE ENSEMBLES OF BOUND ATOMS

The nature of the order-disorder phase transition for an ensemble of bound atoms may be understood on the basis of the lattice model (see, e.g., [6, 59, 60]). Within this model, atoms are located at sites of a crystal lattice and interaction occurs only between nearest neighbors. Then the ordered state is a compact distribution of atoms, which leads to a maximum number of bonds between nearest-neighbor atoms, and the disordered state with a random distribution of atoms corresponds to a maximum entropy and to a loss of some of the bonds between nearest neighbors that occur in the ordered state. The phase transition between these states proceeds by a stepwise change of the total atomic binding energy and the entropy of the evolving distribution. This order-disorder phase transition models the solid-liquid phase transition for an ensemble of bound atoms, with the ordered state being analogous to the solid state and the disordered state analogous to the liquid state. Because this phase transition involves a change of the atomic configuration, the passage to the disordered state occurs by configurational excitation of the system. One can see that the lattice model for configurational excitation of such a system is a simplified void model, in which a void is considered equivalent to a vacancy and additional assumptions are used. Thus, the conclusions following from the lattice model generally apply also to the void model of configuration excitation.

In considering a phase change of clusters, we invoke a certain hierarchy of times for establishment of the various equilibria along the path to the overall transition. We now exhibit the validity of the hierarchy of times in clusters that was the basis of the argument. We assume that a typical time to establish thermal (vibrational) equilibrium in clusters, τ_0, is short compared to a typical time of transition between aggregate states. Roughly, we have $\tau_0 \sim 1/\omega_D$, where ω_D is the Debye frequency. To estimate the time needed for an atom promoted to the cluster surface to move from one position to another or to its ground-state location, we assume for simplicity that all atom positions are separated by identical barriers. Moreover, we take the barrier height to be $U_0 = 0.56D$, as it is at zero temperature (see Fig. 3), even though a temperature increase leads to a decrease of these barriers. A transition of an atom in a three-contact hollow on the cluster surface may proceed in one of the three directions through a triangular transition state, joined there to two nearest neighbors. With τ denoting the transition time to a neighboring position on the cluster surface, we find the mean transition time to the ground state to be 9τ. Thus, under these assumptions, we infer that transitions between different configurational states proceed faster by an order of magnitude than a transition from the surface to a vacancy in the outer shell, normally occupied in the solid state. This allows us to join all the different configurational excitations for promotion of one atom from the outermost shell onto the cluster surface in one liquid aggregate state.

We now estimate the time τ of transition between neighboring positions on the cluster surface on the basis of the Frenkel model [61], according to which the transition occurs if the atomic oscillation energy exceeds the barrier height. Assuming for simplicity that the barrier has an axial symmetry, we take it in the form

$$U(\rho) = U_0 + \frac{1}{2}\frac{d^2U}{d\rho^2}\rho^2,$$

where ρ is the distance from the point of the barrier minimum in the saddle plane. From this, we find the transition rate

$$\begin{aligned}\nu &= \frac{1}{\tau_0}\frac{1}{4\pi R^2}\int_0^\infty 2\pi\rho d\rho \exp\left[-\frac{U(\rho)}{T}\right] \\ &= \frac{T}{2R^2\frac{d^2U}{d\rho^2}}\exp\left(-\frac{U_0}{T}\right),\end{aligned} \tag{31}$$

where T is the current temperature and R is the distance between an atom and the saddle point. Taking

$$\frac{d^2U}{d\rho^2} \sim \frac{D}{R^2}$$

for an estimate, we obtain

$$\tau = \frac{1}{\nu} \sim 50\tau_0 \tag{32}$$

at the melting point $T_m = 0.29D$. Thus, the hierarchy of times is as we used above.

Consequently, we have three typical times that are of importance near the melting point, where the rates of the phase transition in both directions are similar. These times are the time τ_0 of equilibrium establishment for transversal degrees of freedom, during which a definite vibrational temperature is established depending on the aggregate state; the typical time τ for transition between neighboring configurationally excited states; and the typical time of transition between aggregate states. The last is the longest one, but during the time τ, an equilibrium is established within the liquid state. This reflects the lability of that state. We note that from the standpoint of void formation, the case of a 13-atom cluster is special because the liquid aggregate state involves formation of only one void in this case. For larger clusters, the time of transition between the aggregate states is increased in comparison with times of transition between neighboring configurationally excited states. We demonstrate this for a 55-atom cluster, assuming that the number of voids $v = 6$ is the optimal one for the liquid state. This means that the total probability for the cluster to be in any of the configurationally excited states with $v = 5$ is less than for $v = 6$, and for $v = 4$ is still less than that for $v = 5$. To reach the solid aggregate state starting from the liquid, this cluster must pass through configurationally excited states with small probabilities of occurrence, and hence, in most histories, the cluster returns to its initial liquid state many times and eventually reaches the other, solid aggregate state. This means that a typical time of transition between aggregate states is very long compared with the time for passage between neighboring configurationally excited states, particularly in the case of large clusters. Thus, the dynamic coexistence of phases in clusters proceeds such that the vibrational temperature is established fast, next an equilibrium is established between configurationally excited states of the liquid phase, and then the phase transition can proceed during longer times.

At zero temperature, the vacancy and the relaxed void become equivalent if we neglect the vacancy-atom interaction. Real parameters of voids take this interaction into account, and hence the relaxation has its effect on the thermal motion of atoms upon configurational excitation. Of course, the lower the temperature, the less is the configurational excitation and the less is the vibrational excitation as well. Evidently, the separation of the configurational excitation from thermal vibrations of atoms that we have used is valid only at low to moderate temperatures, and is better for clusters with completed outer shells than for others. Therefore, we use the void concept primarily for clusters with complete shells, such as those consisting of 7, 13, 19, 55, 147, … atoms. In these cases, there is a solid-liquid coexistence region of temperature and pressure within which the probability distribution of the total kinetic energy is distinctly bimodal [11, 26] for an isolated cluster. The occurrence and persistence of these two aggregate states allows us to use the approach of two aggregate states [44], which is an analogue of the solid and liquid aggregate states for bulk systems. We note that in reality, several types of configurational excitations can be observed that correspond to excitation of different cluster shells [27, 28].

Next, for some clusters with incomplete outer shells, thermodynamically stable states of configurational excitation are absent, in particular, for clusters consisting of 8 and 14 atoms [26], because only a small entropy (and free energy) jump separates the states; this is much the same situation as occurs with excited states of atoms with open shells. Therefore, the real behavior of excitations of open-shell clusters with pair interactions may be more complicated than that within the framework of the void model of a cluster with two aggregate states. Nevertheless, this model is useful for understanding and description of the clusters with a pairwise atomic interaction.

Parameters of melting for atomic clusters and macroscopic inert gases

	LJ_{13}	LJ_{55}	bulk inert gases
T_m/D	0.29	0.31	0.58
$\Delta E/D$	2.5	16 ± 1	0.98n
$(T_{sol} - T_{liq})/T_m$	0.22	0.31 ± 0.02	0.56
ΔS_0	5.2	31 ± 2	0.73n
ΔS_m	8.6	48 ± 4	1.68n
$\Delta S_0/\Delta S_m$, %	60	65 ± 10	44

The table presents some parameters for the liquid states of Lennard-Jones clusters consisting of 13 and 55 atoms, which we obtained from the analysis of the results of molecular dynamics computer simulations. In this table, LJ_n is a cluster consisting of n atoms with the Lennard-Jones interaction potential, D is the depth of the potential well, T_m is the melting point, ΔE is the phase transition energy for an isolated cluster at the melting point, T_{sol} and T_{liq} are the effective (kinetic-energy-based) temperatures of the solid and liquid states for an isolated cluster at the melting point, and ΔS_0 and ΔS_m are the respective entropy jumps for the phase transition at zero temperature and the melting point. In determining the difference $T_{sol}–T_{liq}$, we assume the heat capacity for each aggregate state to be given by the Dulong–Petit law.

In the table, we also include the parameters of bulk inert gases consisting of n atoms, which were found [15–17, 19] on the basis of measured parameters of condensed inert gases and are averaged over classical inert gases (Ne, Ar, Kr, Xe). Then the reduced entropy jump in inert gases near the triple point is

$$\Delta S_m/n = 1.68 \pm 0.03$$

[18, 43, 62]. Considering the entropy jump at zero temperature as a result of vacancy formation in a solid, we then obtain

$$\Delta S_0 = \ln C_{n+v}^{v} = n\ln\frac{n+v}{n} + v\ln\frac{n+v}{v}, \qquad (33)$$

where n is the number of atoms in the system, v is the number of vacancies, and C_{n+v}^{v} is the number of ways to remove v atoms from the initial lattice containing $n + v$ atoms. For condensed inert gases [15–17, 19], we have

$$n/v = 3.12 \pm 0.01,$$

which gives

$$\Delta S_0/n = 0.73$$

as included in the table.

It follows from the data in the table that in all the cases under consideration, the atomic thermal motion makes a very large contribution to the entropy jump at the melting point. This effect is very important because it holds down the temperature of the phase transition or even makes it possible at all, in principle. In addition, the thermal contribution to the entropy jump can solve the paradox of the phase transition, which we now consider. In practice, it is convenient to use the Lindemann criterion [29, 30] for the melting point of an ensemble of bound atoms. According to this criterion, melting starts if the ratio of the amplitude of atom oscillations to the distance between nearest neighbors reaches a value in the range 0.10–0.15. With computer modeling of clusters, more precise criteria of the phase transition were introduced, using the correlations in positions of two cluster atoms. In particular, this correlation function can use the Etters–Kaelberer parameter [31–33] or the Berry parameter [12, 34]. These parameters are proportional to the mean fluctuation of the distance between two atoms, which, similarly to the Lindemann index, falls into different ranges for the solid and liquid states; this difference allows us to distinguish a cluster's state. These parameters are connected with thermal motion of atoms, while the melting, i.e., the lability of the liquid, results from configurational excitation of an ensemble of bound atoms.

One can see an apparent contradiction between the nature of the phase transition that we attribute to configurational excitation and the practical criterion signifying this transition, which is based on thermal motion of atoms. This contradiction disappears when we account for the influence of thermal excitation on the entropy of this transition in accordance with formula (21). The second part of this formula accounts for the apparent paradox of the thermal motion in the entropy jump, and hence, if we understand the origin of the paradox, the amplitude-based criteria of the phase transition become natural. The nature of this term results from the lower density of atoms in the liquid state and from the associated larger entropy of the atomic vibrations. Naturally, the entropy jump due to vibrations of atoms increases with increasing the temperature. Thus, although the method of calculation separates configurational excitation from the thermal motion of the bound atoms, the latter gives a contribution to the entropy change of the transition.

We conclude that because thermal motion of atoms gives a significant contribution to the entropy jump, this effect improves conditions of the phase transition or can even be a required condition for the phase transition. When we consider a bulk system of bound atoms, we base our argument on the model in which the liquid state is formed from the solid state by removal of internal atoms. Then the system relaxes to the liquid state by shrinking, such that vacancies of the crystal lattice are transformed into voids. The entropy of this configurational excitation follows from this intermediate state with vacancies, and the void concept [14] describes the phase transition. Using the void concept for the analysis of the phase transitions allows one to understand its nature more deeply.

6. CONCLUSIONS

The void concept for configurational excitation of ensembles of bound atoms is the basis for their analysis. This concept follows from a general approach of local minima of the potential energy surface for an ensemble of interacting atoms [20]; a simplified version of this approach allows us to analyze the results of cluster computer simulations by molecular dynamics. On the basis of this analysis, we can understand some aspects of the behavior of ensembles of bound atoms at the phase transition in detail. In particular, there is a difference in the transition parameters for an isolated cluster at constant energy and a similar cluster in a thermostat. The entropy jump of the phase transition includes two contributions, both of which are important: the thermal, vibrational motion of atoms (because the solid state is characterized by a more compact distribution and correspondingly by a lower entropy than the liquid aggregate state at this temperature) and the configurational excitation that introduces the voids, providing the basis of the fluidity of the liquid. The thermal effect in the entropy jump at the phase transition provides the validity of melting criteria based on thermal motion of atoms, whereas the "nature" of the phase transition consists in the configurational excitation.

The void concept for configurational excitation of ensembles of bound atoms, interpreted with the help of the results of computer simulations and thermodynamic parameters of condensed inert gases, gives a deepened understanding of the phase transition in these ensembles.

ACKNOWLEDGMENTS

This paper is supported in part by the Russian Foundation for Basic Research (project no. 03-02-16059). R. S. B. wishes to acknowledge the support of a Grant from the National Science Foundation.

REFERENCES

1. O. Echt, K. Satler, and E. Recknagel, Phys. Rev. Lett. **47**, 1121 (1981).
2. O. Echt, A. Reyes Flotte, M. Knapp, *et al.*, Ber. Bunsenges. Phys. Chem. **86**, 860 (1982).
3. J. Xie, J. A. Northby, D. L. Freemann, and J. P. Doll, J. Chem. Phys. **91**, 612 (1989).
4. K. Huang, *Statistical Mechanics* (Wiley, New York, 1963).
5. R. H. Brout, *Phase Transitions* (Benjamin, New York, 1965).
6. R. Kubo, *Statistical Mechanics: An Advanced Course with Problems and Solutions* (North-Holland, Amsterdam, 1965).
7. D. ter Haar and H. Wergeland, *Elements of Thermodynamics* (Addison-Wesley, New York, 1966).
8. R. P. Feynman, *Statistical Mechanics: A Set of Lectures* (Benjamin, Reading, Mass., 1972; Mir, Moscow, 1975).

9. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics*, 3rd ed. (Nauka, Moscow, 1976; Pergamon, Oxford, 1980), Part 1.
10. R. S. Berry, J. Jellinek, and G. Natanson, Phys. Rev. A **30**, 919 (1984).
11. J. Jellinek, T. L. Beck, and R. S. Berry, J. Chem. Phys. **84**, 2783 (1986).
12. R. S. Berry, T. L. Beck, H. L. Davis, and J. Jellinek, Adv. Chem. Phys. **90**, 75 (1988).
13. R. S. Berry, in *Theory of Atomic and Molecular Clusters*, Ed. by J. Jellinek (Springer, Berlin, 1999), p. 1.
14. H. Reiss, H. L. Frisch, and J. L. Lebowitz, J. Chem. Phys. **31**, 369 (1959).
15. B. M. Smirnov, Zh. Éksp. Teor. Fiz. **112**, 1847 (1997) [JETP **85**, 1010 (1997)].
16. B. M. Smirnov, Inorg. Mater. **35**, 562 (1999).
17. B. M. Smirnov, in *Nucleation Theory and Applications*, Ed. by J. W. P. Schmelzer, G. Röpke, and V. B. Priezzhev (JINR, Dubna, 1999), p. 355.
18. B. M. Smirnov, *Clusters and Small Particles in Gases and Plasmas* (Springer, New York, 1999).
19. R. S. Berry and B. M. Smirnov, in *Nucleation Theory and Applications*, Ed. by J. W. P. Schmelzer, G. Röpke, and V. B. Priezzhev (JINR, Dubna, 2002), p. 340.
20. D. Wales, *Energy Landscapes* (Cambridge Univ. Press, Cambridge, 2003).
21. D. S. Corti, P. G. Debenedetti, S. Sastry, and F. H. Stillinger, Phys. Rev. E **55**, 5522 (1997).
22. K. D. Ball and R. S. Berry, J. Chem. Phys. **111**, 2060 (1999).
23. T. Komatsuzaki and R. S. Berry, J. Chem. Phys. **110**, 9160 (1999).
24. D. J. Wales, J. P. K. Doye, M. A. Miller, *et al.*, Adv. Chem. Phys. **115**, 1 (2000).
25. B. Vekhter and R. S. Berry, J. Chem. Phys. **106**, 4644 (1997).
26. T. L. Beck, J. Jellinek, and R. S. Berry, J. Chem. Phys. **87**, 545 (1987).
27. R. E. Kunz and R. S. Berry, Phys. Rev. Lett. **71**, 3987 (1993).
28. R. E. Kunz and R. S. Berry, Phys. Rev. E **49**, 1895 (1994).
29. F. A. Lindemann, Z. Phys. **11**, 609 (1910).
30. J. P. Hansen and L. Verlet, Phys. Rev. **184**, 151 (1969).
31. R. D. Etters and J. B. Kaelberer, Phys. Rev. A **11**, 1068 (1975).
32. R. D. Etters and J. B. Kaelberer, J. Chem. Phys. **66**, 5112 (1977).
33. J. B. Kaelberer and R. D. Etters, J. Chem. Phys. **66**, 3233 (1977).
34. Y. Zhou, M. Karplus, K. D. Ball, and R. S. Berry, J. Chem. Phys. **116**, 2323 (2002).
35. M. R. Hoare and P. Pal, Adv. Phys. **20**, 161 (1971).
36. M. R. Hoare and P. Pal, Adv. Phys. **24**, 645 (1975).
37. M. R. Hoare, Adv. Chem. Phys. **40**, 49 (1979).
38. F. H. Stillinger and T. A. Weber, Phys. Rev. A **25**, 978 (1982).
39. F. H. Stillinger and T. A. Weber, Phys. Rev. A **28**, 2408 (1983).
40. D. J. Wales and R. S. Berry, J. Chem. Phys. **92**, 4283 (1990).
41. D. ter Haar, *Elements of Thermostatics*, 2nd ed. (Addison-Wesley, New York, 1966).
42. B. M. Smirnov, *Clusters and Small Particles in Gases and Plasmas* (Springer, New York, 1999).
43. B. M. Smirnov, Phys. Usp. **37**, 1079 (1994).
44. R. S. Berry and B. M. Smirnov, J. Chem. Phys. **114**, 6816 (2001).
45. R. S. Berry, submitted to Isr. J. Chem. (2004).
46. M. Schmidt, R. Kusche, W. Krommüller, *et al.*, Phys. Rev. Lett. **79**, 99 (1997).
47. M. Schmidt, R. Kusche, B. von Issendorf, *et al.*, Nature **393**, 238 (1998).
48. M. Schmidt, R. Kusche, T. Hippler, *et al.*, Phys. Rev. Lett. **86**, 1191 (2001).
49. M. Schmidt, T. Hippler, J. Donges, *et al.*, Phys. Rev. Lett. **87**, 203402 (2001).
50. G. Natanson, F. Amar, and R. S. Berry, J. Chem. Phys. **78**, 399 (1983).
51. J. Jellinek, T. L. Beck, and R. S. Berry, Chem. Phys. Lett. **107**, 227 (1984).
52. R. S. Berry, Chem. Rev. **93**, 2379 (1993).
53. R. S. Berry and B. M. Smirnov, Zh. Éksp. Teor. Fiz. **120**, 889 (2001) [JETP **93**, 777 (2001)].
54. R. S. Berry and B. M. Smirnov, Zh. Éksp. Teor. Fiz. **125**, 414 (2004) [JETP **98**, 366 (2004)].
55. P. Labastie and R. L. Whetten, Phys. Rev. Lett. **65**, 1567 (1990).
56. H. P. Cheng and R. S. Berry, Phys. Rev. A **45**, 7969 (1992).
57. J. A. Northby, J. Chem. Phys. **87**, 6166 (1987).
58. D. J. Wales, Chem. Phys. Lett. **166**, 419 (1990).
59. W. Bragg and H. J. Williams, Proc. R. Soc. London, Ser. A **145**, 699 (1934).
60. J. M. Ziman, *Models of Disorder: The Theoretical Physics of Homogeneously Disordered Systems* (Cambridge Univ. Press, Cambridge, 1979; Mir, Moscow, 1982).
61. Ya. I. Frenkel, *Kinetic Theory of Liquids* (Akad. Nauk SSSR, Moscow, 1946; Oxford Univ. Press, Oxford, 1946).
62. B. M. Smirnov, Phys. Usp. **44**, 1229 (2001).

Journal of Experimental and Theoretical Physics, Vol. 100, No. 6, 2005, pp. 1142–1152.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 6, 2005, pp. 1297–1309.
Original Russian Text Copyright © 2005 by Karkin, Naumov, Goshchitskii, Balbashov.

ELECTRONIC PROPERTIES OF SOLIDS

Galvanomagnetic Properties of Atomically Disordered Sr_2RuO_4 Single Crystals

A. E. Karkin[a], S. V. Naumov[a], B. N. Goshchitskii[a], and A. M. Balbashov[b]
[a]*Institute of Metal Physics, Ural Division, Russian Academy of Sciences, Yekaterinburg, 620219 Russia*
[b]*Moscow Power Engineering Institute, Moscow, 111250 Russia*
e-mail: karkin@uraltc.ru

Received September 9, 2004

Abstract—The effect of neutron-bombardment-induced atomic disorder on the galvanomagnetic properties of Sr_2RuO_4 single crystals has been experimentally studied in a broad range of temperatures (1.7–380 K) and magnetic fields (up to 13.6 T). The disorder leads to the appearance of negative temperature coefficients for both the in-plane electric resistivity (ρ_a) and that along the c axis (ρ_c), as well as the negative magnetoresistance $\Delta\rho$, which is strongly anisotropic to the magnetic field orientation ($\mathbf{H} \parallel \mathbf{a}$ and $\mathbf{H} \parallel \mathbf{c}$), with the easy magnetization direction along the c axis and a weak dependence on the probing current direction in the low-temperature region. The experimental $\rho_a(T)$ and $\rho_c(T)$ curves obtained for the initial and radiation-disordered samples can be described within the framework of a theoretical model with two conductivity channels. The first channel corresponds to the charge carriers with increased effective masses ($\sim 10m_e$, where m_e is the electron mass) and predominantly electron–electron scattering, which leads to the quadratic temperature dependences of ρ_a and ρ_c. The second channel corresponds to the charge carriers with lower effective masses exhibiting magnetic scattering at low temperatures, which leads to the temperature dependence of the $\rho_{a,\,c}(T) \propto 1/T$ type. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The proper understanding of the mechanism of high-temperature superconductivity in cuprates requires a theory adequately describing their properties, primarily in the normal (metallic) state. The electron transport in cuprate high-temperature superconductors (HTSCs) exhibits a number of anomalous features not inherent in usual metals. This circumstance hinders description of the properties of such HTSCs within the framework of a usual model of the Fermi liquid. While the temperature dependence of the in-plane electric resistivity in HTSCs has a "metallic" shape (with $d\rho_a/dT > 0$), the behavior of the out-of-plane component (in c axis direction) is more like that in semiconductors (with $d\rho_c/dT < 0$) [1, 2]. In addition, ρ_a exhibits a weaker (approximately linear) dependence on the temperature as compared to behavior of the $\rho_a(T) \propto T^2$ type expected for compounds with strong electron correlations. Other unexpected features are (i) a strong temperature dependence of the Hall coefficient [3] and (ii) the negative values of magnetoresistance ($\Delta\rho < 0$) at $T > T_C$ [4, 5].

In order to develop a theory adequately describing the properties of HTSC compounds in the normal state, it is necessary to base the analysis on reliable experimental data on the galvanomagnetic properties in the low-temperature limit ($T \longrightarrow 0$), which would provide reliable information on the features of electron states near the Fermi level. However, rather high values of the superconducting transition temperature ($T_C \sim 100$ K) and the second (upper) critical field ($H_{C2} \sim 100$ T) make such experiments very difficult. The "anomalous" properties of HTSCs are by no means unique, since similar features in the transport properties were observed in many other systems with strong electron correlations such as compounds with heavy fermions [6] or Sr_2RuO_4 [7, 8]. However, the mechanisms underlying this behavior may be different. In particular, a nonmetallic behavior of the electric resistivity along the c axis in HTSC-like compounds is usually related to the hopping mechanism of conductivity [9], which seems to be quite well justified for these virtually quasi-two-dimensional systems. In contrast, such a behavior in almost isotropic systems with heavy fermions is attributed to a magnetic scattering of the Kondo type [6].

The systems with strong electron correlations are extremely sensitive to isovalent doping, as well as to other kinds of atomic disorder, which makes it possible to use the methods of disordering for studying the features of electron spectrum near the Fermi level. In contrast to the case of usual metals, the atomic disorder in HTSCs induced, in particular, by fast neutron bombardment, leads to a cardinal rearrangement of the electron spectrum and results in electronic phase transitions of the metal–dielectric type [10, 11]. For example, a "semiconductor" character of the in-plane conductivity arises when the sample is still in superconducting state [12]. It should be noted that radiation-induced disorder was used [10–12] as a means of directional modification of the electron properties in order to reveal "anomalous" features of electron transport in the initial ordered state. Obvious advantages of this method are the homogeneous distribution of defects over the sample volume at an unchanged stoichiometric composition. The radiation-induced disorder is reversible: the resulting nonequilibrium disordered state can be gradually shifted toward a higher order by means of sequential high-temperature anneals until compete restoration of the initial (ordered) state [13].

Strontium ruthenium oxide Sr_2RuO_4 is frequently considered as an electronic analog of HTSC compounds. According to published data, high quality of Sr_2RuO_4 single crystals and low values of their superconducting transition temperature ($T_C < 2$ K) and the upper critical field ($H_{C2} < 0.1$ T) made it possible to determine the main band structure parameters and the Fermi surface topology of this compound using powerful experimental methods such as the de Haas–Van Alphen effect [14, 15], cyclotron resonance [16], angular-resolved photoemission [17], and some other [18–23] and numerical band-structure calculations [24], so that now these properties are much more reliably established for Sr_2RuO_4 than for any HTSC compound. The Fermi surface of Sr_2RuO_4 consists of one hole cylinder (α sheet) and two electron cylinders (β and γ sheets) with the corresponding wavevectors $k_F = 0.3$, 0.62, and 0.75 $Å^{-1}$ and effective electron masses 3.4, 6.6, and 12.0 (in m_e units), respectively. The main contribution to the conductivity and the Hall effect is due to the γ surface corresponding to heavier electrons [14].

In the initial ordered Sr_2RuO_4, the behavior of $\rho_a(T)$ and $\rho_c(T)$ at low temperatures ($T < 30$ K) is described by a function of the $\rho \propto T^2$ type, which is quite natural for the Fermi liquid. In the region of higher temperatures ($T > 100$ K), the behavior of the conductivity qualitatively changes: $\rho_a(T)$ grows slower than in the low-temperature region and can be described, as demonstrated by Mackenzie *et al.* [25], by the sum of the quadratic and linear terms; $\rho_c(T)$ passes through a maximum at $T \sim 100$ K and then decreases ($d\rho_c/dT < 0$) in the interval of temperatures up to ~600 K [26]. It was shown [27] that the out-of-plane conductivity can be also described in a broad temperature range by a sum of two terms, one of which is proportional to the in-plane conductivity $(\rho_a)^{-1}$ and the other is proportional to the temperature. Because of this qualitatively different behavior of $\rho_a(T)$ and $\rho_c(T)$, the anisotropy of the electric resistivity defined as the ratio ρ_c/ρ_a decreases from about 2000 (at low temperatures) to ~20 at $T \approx 1300$ K [27]. The Hall coefficient R_H strongly depends on the temperature and is negative at $T < 30$ K, positive at $30\text{ K} < T < 130$ K, and negative again at $T > 130$ K [25, 28]. The magnetoresistance of Sr_2RuO_4 exhibits rather involved field and temperature dependences, changing sign from positive (at low T) to negative (at $T \approx 80$ K); in addition, the negative magnetoresistance significantly increases in magnitude for the field H oriented along the c axis [27, 28].

This behavior of galvanomagnetic properties in Sr_2RuO_4, as well as in other HTSC-like compounds is frequently referred to as the transition from a coherent mechanism of conductivity at low T to a noncoherent mechanism at high temperatures. An analogous approach was also used for description of the charge transport in systems with heavy fermions (the transition from a Kondo-lattice at low T to a disordered Kondo system at high temperatures). In the case of strongly anisotropic HTSC-like compounds, this implies that the character of the out-of-plane transport changes from strongly anisotropic metallic to hopping (related to single-particle tunneling induced by magnons or other appropriate quasi-particles. Recently, Minakata and Maeno [29] studied the effect of replacement of ruthenium ions by isovalent nonmagnetic titanium ions (with respect to the influence on the electron properties, this kind of doping is closest to the radiation-induced disorder used in our study). It was demonstrated that an increase in the content of titanium in $Sr_2RuO_{1-x}Ti_xO_4$ not only results in a gradual decrease in the crossover temperature separating the regions with $d\rho_c/dT > 0$ and $d\rho_c/dT < 0$ (until complete vanishing of the crossover at $x \approx 0.05$), but also leads to the appearance of a low-temperature region where $d\rho_a/dT < 0$ in combination with a quite "metallic" resistivity $\rho_a < 100$ μΩ cm. Evidently, this behavior cannot be interpreted as a kind of noncoherent (hopping) conductivity, at least to a homogeneous electronic system. Thus, the electron transport in Sr_2RuO_4, as well in other HTSC-like compounds, is still not given an adequate description.

2. EXPERIMENT

We have studied the temperature dependence of the electric resistivity of Sr_2RuO_4 single crystals in magnetic fields up to 13.6 T. The measurements were performed on the crystals irradiated by fast neutrons and then subjected to isochronous (20 min) anneals in air at temperatures T_{ann} in the range from 100 to 1000°C. The Sr_2RuO_4 single crystals were obtained by RF crucibleless zone melting technique in air, proceeding from cylindrical ceramic samples prepared by means of a conventional solid-state reaction. The experiments were performed on crystals with mirror-smooth surfaces and typical dimensions 1.5×0.5 mm^2 (in plane) at a thickness of 50–100 μm. The samples were irradiated by fast neutrons to a total fluence of $\Phi = (0.5, 1.0, 1.5, 3, \text{and } 5) \times 10^{19}$ cm^{-2} (two samples for each dose) in a nuclear reactor of the IVV-2M type (the sample temperature during exposure did not exceed 50°C).

The resistivity components ρ_a and ρ_c were determined by the standard four-point-probe Montgomery technique [30] in a temperature range from 1.7 to 380 K. The longitudinal and transverse magnetoresis-

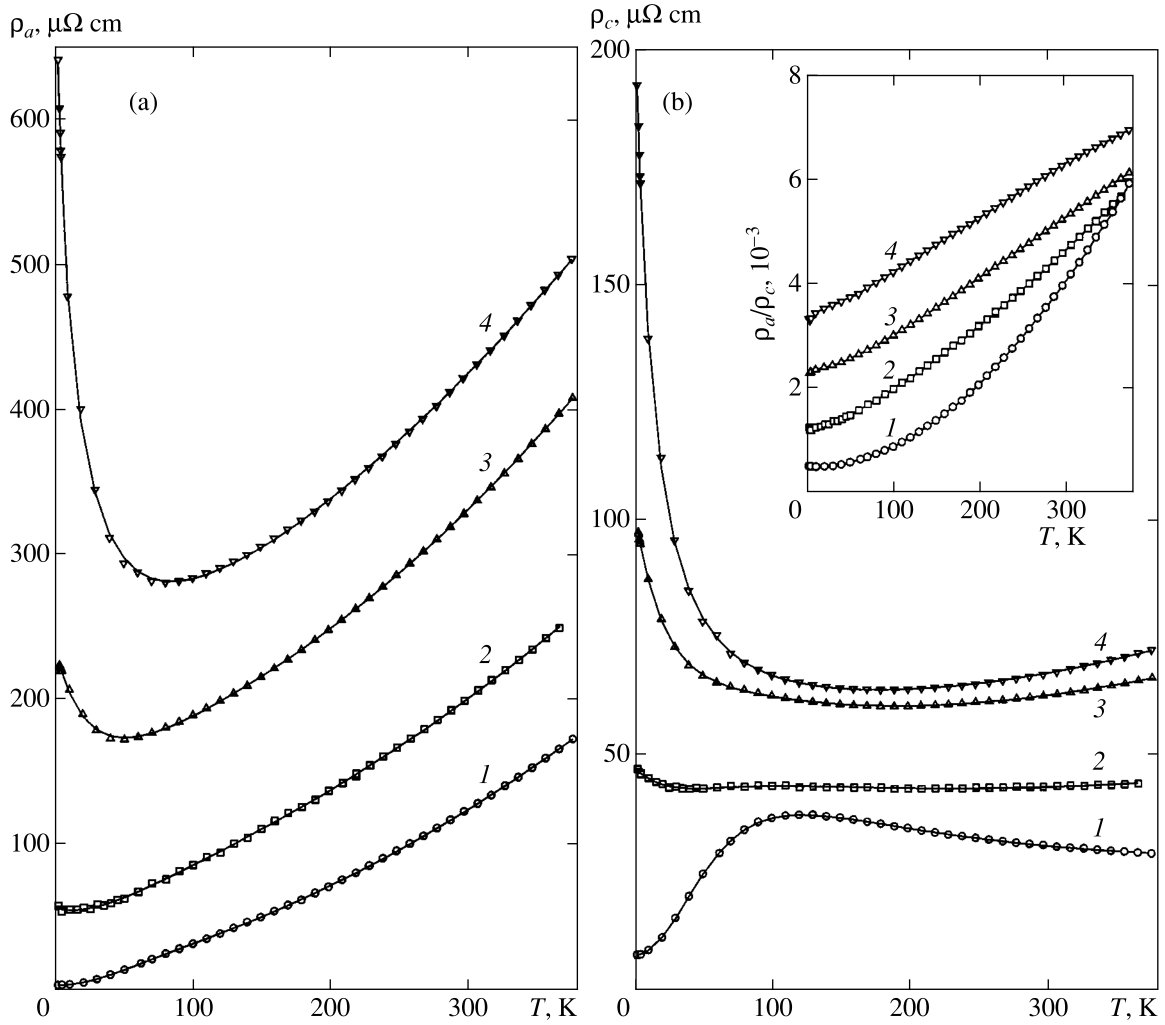


Fig. 1. The temperature dependences of the (a) in-plane (ρ_a) and (b) out-of-plane (ρ_c) resistivities for (*1*) the initial (unirradiated) Sr_2RuO_4 crystal and (*2–4*) the samples irradiated with fast neutrons to various fluences $\Phi = (1.5, 3, \text{and } 5) \times 10^{19}$ cm^{-2}, respectively. The inset shows plots of the anisotropy ρ_a/ρ_c versus temperature. Points represent experimental data, curves show the results of fitting using formulas (3).

tance components were measured at $T = 4.2$ K in magnetic fields up to 13.6 T (with changing directions of the probing current and magnetic field H and switching between the current and potential leads). The results of measurements for each pair of samples irradiated to the same fluence were virtually coinciding. The Montgomery technique gives good results only for the optimum ratio $d_a/d_c \approx (\rho_c/\rho_a)^{1/2}$ of the distances d_a and d_c between electrodes in the a and c directions [30]. A significant variation of the ρ_c/ρ_a ratio as a result of the temperature variation and/or the radiation-induced disordering led to a significant decrease in the accuracy of measurements at lower (higher) temperatures, where the d_a/d_c ratio was above (below) the optimum value. For this reason, below we will analyze in detail the results obtained predominantly for a sample with nearly optimum dimensions ($d_a = 1.85$ mm, $d_c = 0.10$ mm) irradiated to a total fluence of $\Phi = 3 \times 10^{19}$ cm^{-2}. For this sample, the error of measurement of the absolute values of ρ_a and ρ_c (determined predominantly by uncertainty of the sample dimensions) was on the order of 10%. The reproducibility of ρ_a and ρ_c measurements was within 0.05–0.2%.

3. EXPERIMENTAL RESULTS

Figure 1 shows the temperature dependences of the in-plane (ρ_a) and out-of-plane (ρ_c) resistivities for the initial Sr_2RuO_4 crystals and the samples irradiated with fast neutrons to $\Phi = (1.5, 3, \text{and } 5) \times 10^{19}$ cm^{-2}. As can be seen, the radiation-induced disorder leads to a qualitative change in the behavior of both $\rho_a(T)$ and $\rho_c(T)$ in the region of low temperatures ($T < 30$ K). Indeed, even upon irradiation to $\Phi = 1.5 \times 10^{19}$ cm^{-2}, this initial part

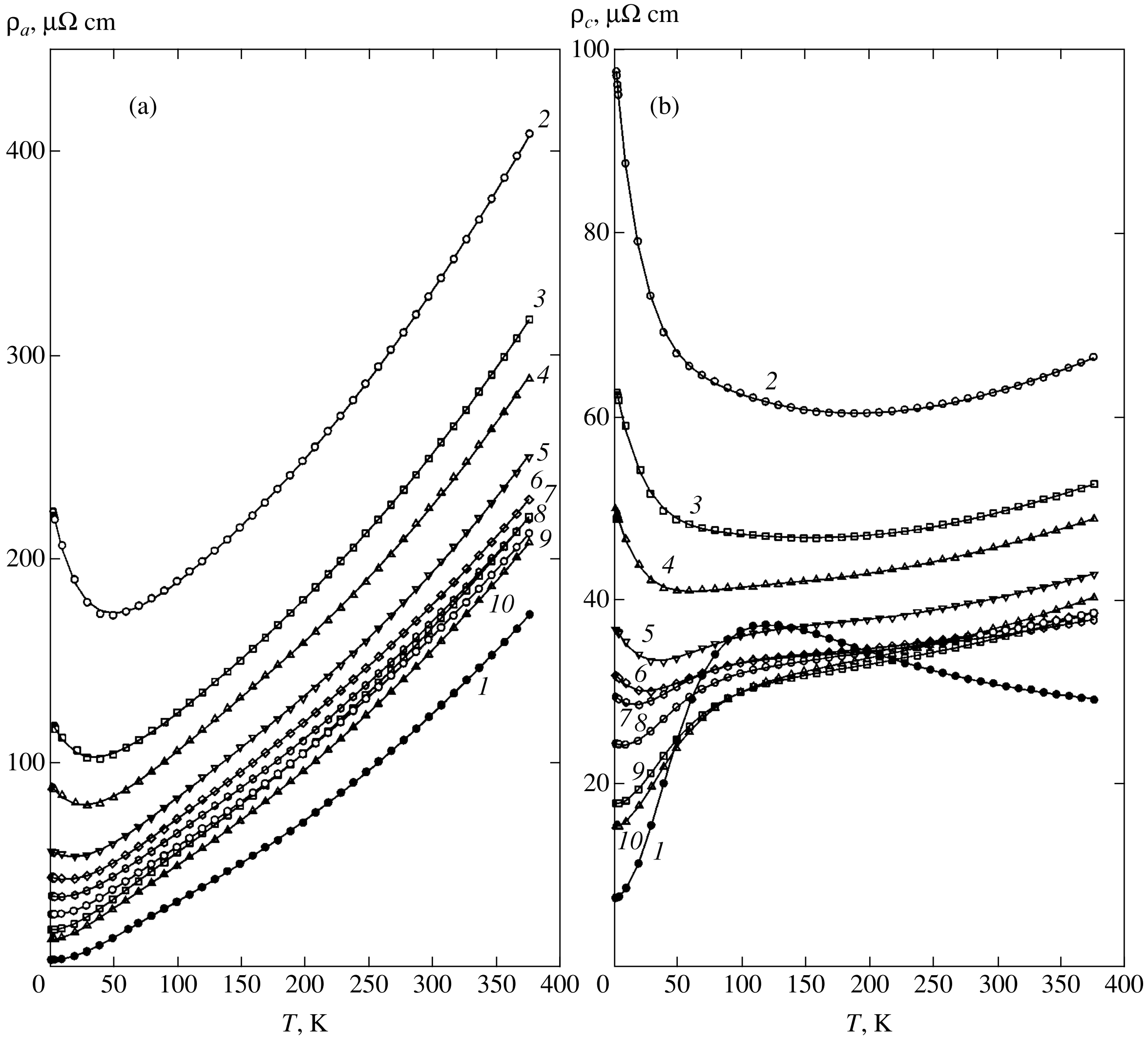


Fig. 2. The temperature dependences of (a) ρ_a and (b) ρ_c for an Sr_2RuO_4 crystal (*1*) in the initial (unirradiated) state, (*2*) upon irradiation with fast neutrons to a fluence of $\Phi = 3 \times 10^{19}$ cm^{-2}, and (*3–10*) after subsequent annealing at 300, 400, 500, 600, 700, 800, 900, and 1000°C, respectively. Points represent experimental data, curves show the results of fitting using formulas (3).

of the curves acquires negative slope ($d\rho/dT < 0$), the magnitude of which rapidly increases with growing disorder. In the region of high temperatures ($T > 100$ K), changes in the behavior of $\rho_a(T)$ and $\rho_c(T)$ are not very significant: as the degree of disorder grows, $d\rho_a/dT$ slightly increases, while $d\rho_c/dT$ changes from a small negative to small positive value. The temperature dependence of the anisotropy (ρ_a/ρ_c) also exhibits significant changes (see the inset in Fig. 1): from almost quadratic dependence for the initial unirradiated) sample to approximately linear dependence for the crystals irradiated to the maximum fluence. These dependences are qualitatively very much like the data reported in [29] for titanium-doped $Sr_2Ru_{1-x}Ti_xO_4$, but the neutron irradiation to $\Phi = 5 \times 10^{19}$ cm^{-2} produces much more pronounced changes in $\rho_a(T)$ and $\rho_c(T)$ than those observed in [29] for a maximum titanium content of $x = 0.09$.

Annealing of the samples irradiated to $\Phi = 3 \times 10^{19}$ cm^{-2} (Fig. 2) at $T_{\text{ann}} = 1000$°C leads to virtually complete restoration of the initial $\rho_a(T)$ curve. At the same time, the $\rho_c(T)$ curves are restored incompletely: while the low-temperature parts of these dependences show a clear tendency to restoration (residual resistivity decreases and the negative slope disappears), the initial high-temperature parts are not restored upon annealing (the slope $d\rho_c/dT$ remains positive and even slightly increases after the treatment at $T_{\text{ann}} = 800$–1000°C).

Figure 3 shows the relative magnetoresistance $\Delta\rho/\rho$ measured at $T = 4.2$ K for the probing current directions in plane ($\Delta\rho_a/\rho_a$) and along the c axis ($\Delta\rho_c/\rho_c$). The measurements were performed for the initial Sr_2RuO_4 crystals and the samples bombarded with fast neutrons

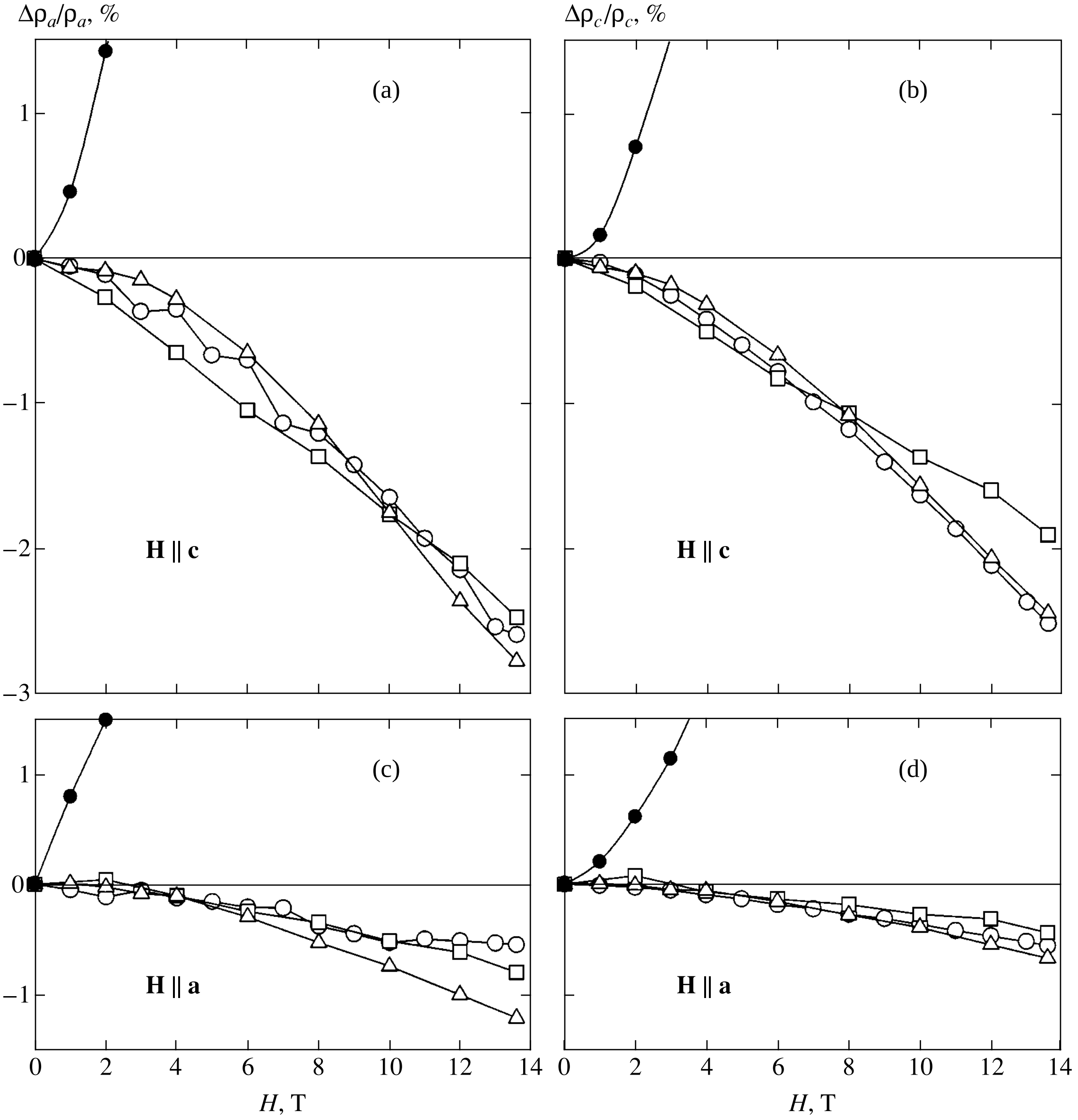


Fig. 3. The relative magnetoresistances (a, c) $\Delta\rho_a/\rho_a$ and (b, d) $\Delta\rho_c/\rho_c$ of Sr_2RuO_4 crystals measured at $T = 4.2$ K for the magnetic field orientations (a, b) **H** || **c** and (c, d) **H** || **a** in the initial state (●) and after irradiation with fast neutrons to various fluences $\Phi = 1 \times 10^{19}$ cm^{-2} (○), 1.5×10^{19} cm^{-2} (□), and 3×10^{19} cm^{-2} (△). Curves connect the points corresponding to the same treatment.

to various degrees of disorder. The residual resistivity ρ_{a0} (which can be considered as a measure of the crystal disorder) varied within broad limits (50–220 μΩ cm). In the initial crystal, all components of the magnetoresistance are positive, in agreement with behavior typical of the usual metals [4, 5, 26]. In disordered samples studied for the field orientation **H** || **c**, the $\Delta\rho_a/\rho_a$ and $\Delta\rho_c/\rho_c$ values are negative in the entire interval of applied fields H and are practically independent of the probing current direction. For **H** || **a**, the magnetoresistance components are much smaller (in absolute value) than those for **H** || **c**, negative for $H > 2$–3 T, and also practically independent of the probing current direction. For both **H** || **a** and **H** || **c** orientations, the $\Delta\rho_a/\rho_a$ and $\Delta\rho_c/\rho_c$ values remain virtually unchanged when the temperature is decreased from 4.2 to 1.5 K.

Figure 4 shows a change in the behavior of $\Delta\rho_c/\rho_c$ in the course of annealing of a sample irradiated to a fluence of $\Phi = 3 \times 10^{19}$ cm^{-2}. After the annealing at T_{ann} = 300–800°C (whereby the residual resistivity ρ_{a0} decreases from 220 to 25 μΩ cm), the $\Delta\rho_c/\rho_c$ value remains almost constant for both **H** || **a** and **H** || **c**. Only the annealing at T_{ann} = 900°C (after which $\rho_{a0} \approx 20$ μΩ cm) leads to the appearance of a significant positive increment in the region of high fields ($H > 6$ T). Not that the initial behavior (as well as that of $\rho_c(T)$, see Fig. 2) is not restored as a result of annealing. Thus, all components of the magnetoresistance—both $\Delta\rho_c/\rho_c$ and $\Delta\rho_a/\rho_a$ (not shown in Fig. 4)—are negative at $T = 4.2$ for $\rho_{a0} > 20$ μΩ cm.

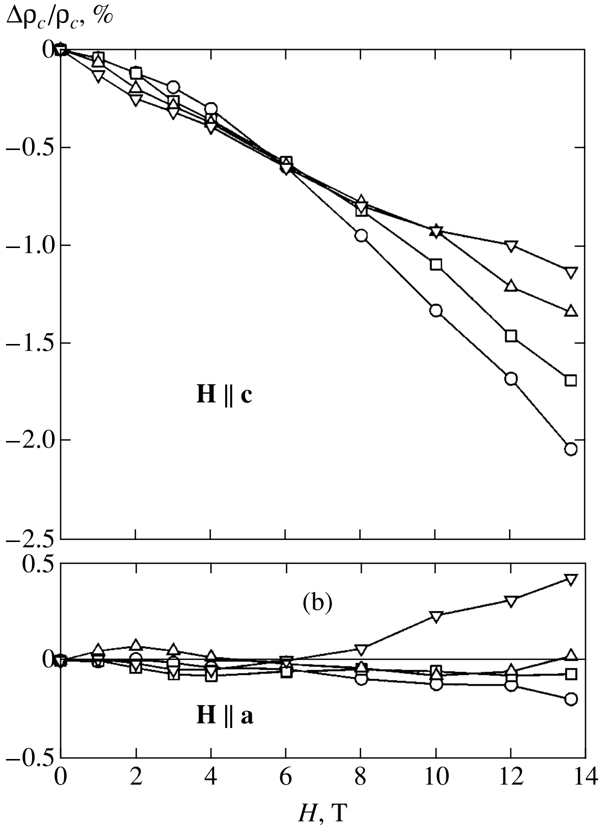


Fig. 4. The relative magnetoresistance $\Delta\rho_c/\rho_c$ of an Sr_2RuO_4 crystal measured at $T = 4.2$ K for the magnetic field orientations (a) **H** || **c** and (b) **H** || **a** upon irradiation with fast neutrons to a fluence of $\Phi = 3 \times 10^{19}$ cm^{-2} and post-annealing at 300 (o), 700 (□), 800 (△), and 900°C (▽). Curves connect the points corresponding to the same treatment.

4. DISCUSSION OF RESULTS

In order to analyze the temperature dependences of ρ_a and ρ_c in (Figs. 1 and 2), let us consider the behavior of derivatives $d\rho_a/dT$ and $d\rho_c/dT$ (Fig. 5). In the range $T > 150$ K, the former derivative can be represented as the sum of a linear function of temperature and a constant, so that $\rho_a(T) = a_0 + a_1T + a_2T^2$. Here, the constant term a_0 rapidly grows with increasing disorder, a_1 rather weakly varies for $\rho_{a0} = 0.3$–0.5 (μΩ cm)/K, and the coefficient a_2 slowly increases from about 0.5 to 1 (nΩ cm)/K^2. By the same token, the behavior of $\rho_c(T)$ in this temperature range can be described as $\rho_c(T) = c_0 + c_1T + c_2T^2$, where the constant term c_0 rapidly increases with growing disorder, while the coefficients c_1 and c_2 virtually remain constant. It should be noted that the values of c_1 and c_2 are much smaller than a_1 and a_2 ($c_1 \approx 0$; $c_2 \approx 0.02$ (μΩ cm)/K^2). In the temperature interval $T < 30$, the behavior of $d\rho_a/dT$ and $d\rho_c/dT$ can also be roughly described as the sum of a constant and linear terms. However, the linear term in this case is much greater than that in the high-temperature, while the constant term decreases from zero for the initial sample to a large negative value for the irradiated crystal. In the intermediate temperature region (30 K < T < 150 K), there is a more (for $d\rho_c/dT$) or less (for $d\rho_a/dT$) pronounced maximum in the temperature dependences of these derivatives.

In order to describe the obtained experimental data, let us represent the conductivity $1/\rho_i(T)$ as the sum of contributions due to two channels denoted by I and II, corresponding to heavy and light charge carriers, respectively:

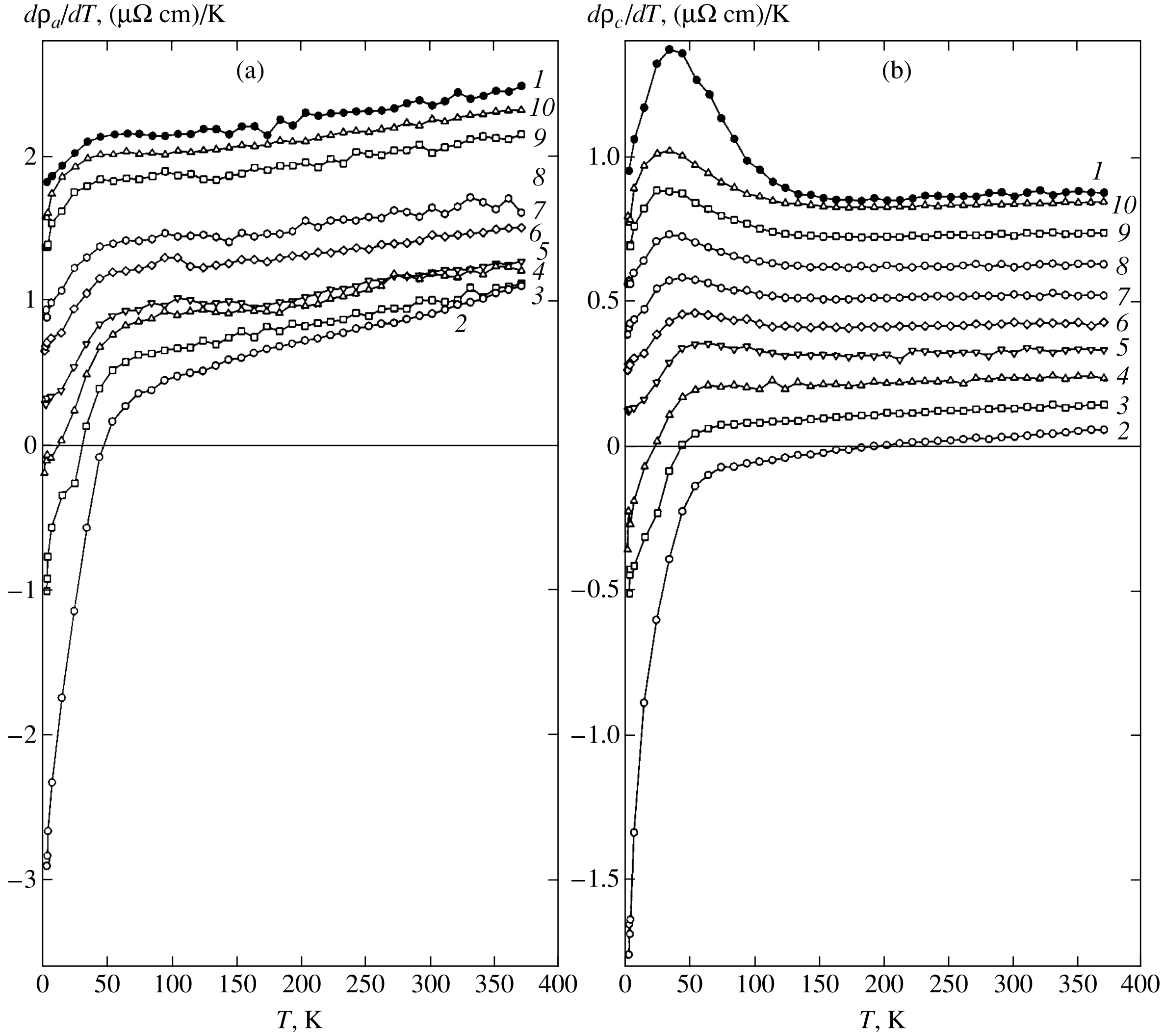
$$\frac{1}{\rho_i(T)} = \frac{1}{\rho_{i\mathrm{I}}(T)} + \frac{1}{\rho_{i\mathrm{II}}(T)}, \quad i = a, c. \tag{1}$$

We assume that $\rho_{i\mathrm{I}}(T) \ll \rho_{i\mathrm{II}}(T)$ at low temperatures and $\rho_{i\mathrm{I}}(T) \gg \rho_{i\mathrm{II}}(T)$ at high temperatures. This implies that $\rho_{i\mathrm{I}}(T)$ and $\rho_{i\mathrm{II}}(T)$ exhibit qualitatively different temperature dependences of the “metallic” and “semiconductor” types, as it was suggested, for example, in [27].

For the conductivity channel I, let us write an expression for the resistivity in the form expected for a metal with strong electron correlations:

$$\rho_{i\mathrm{I}}(T) = A^{\mathrm{I}}_{i0} + A^{\mathrm{I}}_{i2}T^2. \tag{2}$$

As was noted above, the second channel cannot be related to a conductivity of the hopping type. We believe that most consistent explanation of the observed behavior (increase in the resistivity with decreasing temperature) can be given in terms of strong magnetic scattering of the charge carriers. In this case, a natural explanation is provided both for the sign (negative) of all magnetoresistance components $\Delta\rho/\rho$ and for their strong (three- to fivefold) difference in various crystallographic directions. According to [29], the magnetic susceptibility of $Sr_2Ru_{1-x}Ti_xO_4$ contains a contribution of the Curie–Weiss type increasing with x, which exhibits a strong (fivefold) anisotropy with respect to the magnetic field orientation (**H || a** against **H || c**). Therefore, the appearance of negative anisotropic magnetoresistance in radiation-disordered Sr_2RuO_4 crystals is quite natural. It should be noted that the magnetoresistance also contains, besides the negative magnetic contribution, the usual positive contribution related to deviation of the electron trajectories in the applied magnetic field. The latter component should predominate in ordered systems at low temperatures, but it rapidly decreases both with growing disorder (i.e., increasing ρ_0) and with increasing temperature. The change in the sign of $\Delta\rho/\rho$ at $T = 4.2$ K in radiation-disordered



Fig, 5. The temperature dependences of the derivatives (a) $d\rho_a/dT$ and (b) $d\rho_c/dT$ for an Sr_2RuO_4 crystal (*1*) in the initial (unirradiated) state, (*2*) upon irradiation with fast neutrons to a fluence of $\Phi = 3 \times 10^{19}$ cm^{-2}, and (*3*–*10*) after subsequent annealing at 300, 400, 500, 600, 700, 800, 900, and 1000°C, respectively. The curves are sequentially shifted relative to each other along the ordinate axis by (a) 0.2 ($\mu\Omega$ cm)/K and (b) 0.1 ($\mu\Omega$ cm)/K. Curves connect the points corresponding to the same treatment.

samples is observed (Fig. 3) for $\rho_0 \approx 15$–20 $\mu\Omega$ cm, while in the ordered crystals it takes place at $T \sim 80$ K ($\rho_{80} \approx 15$ $\mu\Omega$ cm) [27].

One possible mechanism of magnetic scattering is offered by the Kondo effect, known to contribute with an additional logarithmic term $\rho_m \sim -\ln T$ in $\rho(T)$. However, it was demonstrated [31] that the logarithmic form of magnetic scattering is valid only for dilute systems, whereas more concentrated ones are characterized by an approximately linear additional low-temperature term $\rho_m \sim -T$, which was repeatedly observed in disordered systems (such as (La–Ce)Cu_2Si_2) with heavy fermions [32]. Our system (see Fig. 1) exhibits a clearly pronounced linear temperature dependence of both $\rho_a(T)$ and $\rho_c(T)$. In a broad temperature range, a more realistic form of the magnetic contribution can be provided by $\rho_m \sim 1/(T + T_0)$, which correctly described both the low-temperature (linear) behavior and the damping at higher temperatures. Any other interpolation with analogous behavior in the low- and high-temperature limits can be used as well.

By analogy with [27], let us attribute the magnetic scattering (for the probing current and magnetic field parallel to the axis *c*) entirely to channel II (it should be noted that the negative magnetoresistance was related in [27] to the Zeeman splitting of localized levels involved in the noncoherent hopping process, rather than to the magnetic scattering). We shall also take into account that, according to the observed behavior of $d\rho_a/dT$ (Fig. 5), the theoretical function $\rho_a(T)$ at high temperatures must include a linear term in addition to the quadratic term.

The final expression, which was used for the description of experimental data (solid curves in Figs. 1 and 2), is as follows:

$$\frac{1}{\rho_i(T)} = \frac{1}{\rho_{i\mathrm{I}}(T)} + \frac{1}{\rho_{i\mathrm{II}}(T)},$$

$$\rho_{i\mathrm{I}}(T) = A_{i0}^{\mathrm{I}} + A_{i2}^{\mathrm{I}}T^2,$$

$$\rho_{i\mathrm{II}}(T) = A_{i0}^{\mathrm{II}} + A_{i1}^{\mathrm{II}} + A_{i2}^{\mathrm{II}}T^2 + \rho_{im}, \qquad (3)$$

$$\rho_{im} = \frac{B_{im}}{T + T_{i0}},$$

$$i = a, c.$$

As can be seen from Figs. 1 and 2, using these formulas and varying the fitting parameters A_{i0}^{I}, A_{i2}^{I}, A_{i0}^{II}, A_{i1}^{II}, A_{i2}^{II}, B_{im}, and T_{i0}, it is possible to describe the experimental curves of $\rho_a(T)$ and $\rho_c(T)$ with high precision for the samples disordered by irradiation to various degrees. The fitting parameters exhibit monotonic variation with the degree of disorder.

Figure 6 shows the plots of $\rho_{a\mathrm{I}}(T)$, $\rho_{c\mathrm{I}}(T)$, $\rho_{a\mathrm{II}}(T)$, and $\rho_{c\mathrm{II}}(T)$ functions obtained by subtracting conductivities calculated for each of the two channels from the experimental $1/\rho_a(T)$ and $1/\rho_c(T)$ curves. For example,

$$\rho_{a\mathrm{I}}(T)$$

$$= \left\{\frac{1}{\rho_a(T)} - \left(A_{a0}^{\mathrm{II}} + A_{a1}^{\mathrm{II}}T + A_{a2}^{\mathrm{II}}T^2 + \frac{B_{am}}{T + T_{a0}}\right)\right\}^{-1},$$

$$\rho_{a\mathrm{II}}(T) = \left\{\frac{1}{\rho_a(T)} - \frac{1}{A_{a0}^{\mathrm{I}} + A_{a2}^{\mathrm{I}}T^2}\right\}^{-1}.$$

Figure 7 presents the magnetic contributions ρ_{am} and ρ_{cm} in conductivity channel II for a sample irradiated to a fluence of 3×10^{19} cm^{-2}. These functions were calculated (similarly to those in Fig. 6) as

$$\rho_{im}(T) = \left\{\frac{1}{\rho_i(T)} - \frac{1}{A_{i0}^{\mathrm{I}} + A_{i2}^{\mathrm{I}}T^2}\right\}^{-1}$$

$$- (A_{i0}^{\mathrm{II}} + A_{i1}^{\mathrm{II}}T + A_{i2}^{\mathrm{II}}T^2),$$

where $\rho_i(T)$ are the experimental temperature dependences, and A_{i0}^{I}, A_{i2}^{I}, A_{i0}^{II}, A_{i1}^{II}, and A_{i2}^{II} are fitting parameters. As can be seen, the magnetic contributions are described with good accuracy in a broad temperature range from 1.7 to 380 K by functions of the type $\rho_{im}(T) \propto T^{-1}$.

For the first channel (see Figs. 6a and 6b plotted on a double logarithmic scale), the quadratic dependence $\rho_{i\mathrm{I}}(T) = A_{i0}^{\mathrm{I}} + A_{i2}^{\mathrm{I}}T^2$ well fits to the experimental data in the entire temperature range, and even for the highly disordered samples at high temperatures, where the relative contribution of $1/\rho_{i\mathrm{I}}(T)$ to the total conductivity is as small as 5–10%. The coefficient at the quadratic term (which is related to the electron–electron scattering) is $A_{c2}^{\mathrm{I}} \approx 10$–15 (μΩ cm)/K^2 and weakly changes with increasing disorder. The residual resistivity A_{c0}^{I} increases from about 8 mΩ cm for the initial crystal to 100 mΩ cm for the irradiated samples. The coefficient A_{a2}^{I} increases from about 5 (nΩ cm)/K^2 for the initial crystal to 25 (nΩ cm)/K^2 for the disordered samples. The residual resistivity A_{a0}^{I} increases from about 4 μΩ cm for the initial crystal to 230 μΩ cm for the irradiated samples. Since channel I gives the main contribution to the total conductivity al low temperatures, it is the stronger growth in A_{a0}^{I} and A_{a2}^{I} in comparison to A_{c0}^{I} and A_{c2}^{I} upon the radiation-induced disordering that predominantly accounts for the increase in the low-temperature anisotropy ρ_a/ρ_c (see the inset in Fig. 1). The observed variation of the parameters A_{i0}^{I} and A_{i2}^{I} for the first conductivity channel in the samples with growing disorder can be interpreted in terms of an increase in the effective electron mass for the in-plane transport (m_a), at a relatively weak variation of the effective electron mass along the c axis (m_c).

The curves of $\rho_{a\mathrm{II}}(T)$ and $\rho_{c\mathrm{II}}(T)$ for the second conductivity channel are presented in Figs. 6c and 6d (on a linear scale). As can be seen, the parameters B_{im}, A_{i0}^{II}, A_{i1}^{II}, and A_{i2}^{II} show relatively small variations depending on the degree of disorder, in agreement with the observed high-temperature behavior of the derivatives $d\rho_a/dT$ and $d\rho_c/dT$ (Fig. 5). The main difference between the transport along the c axis and the in-plane transport in channel II is the relatively small value of A_{c2}^{II}. For this reason, the $A_{c2}^{\mathrm{I}}/A_{c2}^{\mathrm{II}}$ ratio is greater by approximately two orders of magnitude than the $A_{a2}^{\mathrm{I}}/A_{a2}^{\mathrm{II}}$ ratio. As a result, the region of the negative slope of $d\rho_c/dT$ extends to lower temperatures. As can be seen from Fig. 5, only $\rho_{c\mathrm{II}}(T)$ for the initial sample deviates from the general trend in variation of the temperature dependence of resistivity, which is related for the most part to smallness of the fitting parameter A_{c2}^{II} in comparison to the corresponding values for irradiated and annealed samples.

The physical meaning of the terms $A_{i1}^{\mathrm{II}}T$ and $A_{i2}^{\mathrm{II}}T^2$ in the second conductivity channel is not as clear. For the in-plane transport, the A_{a2}^{II} value is probably related to the process of electron–electron scattering involving

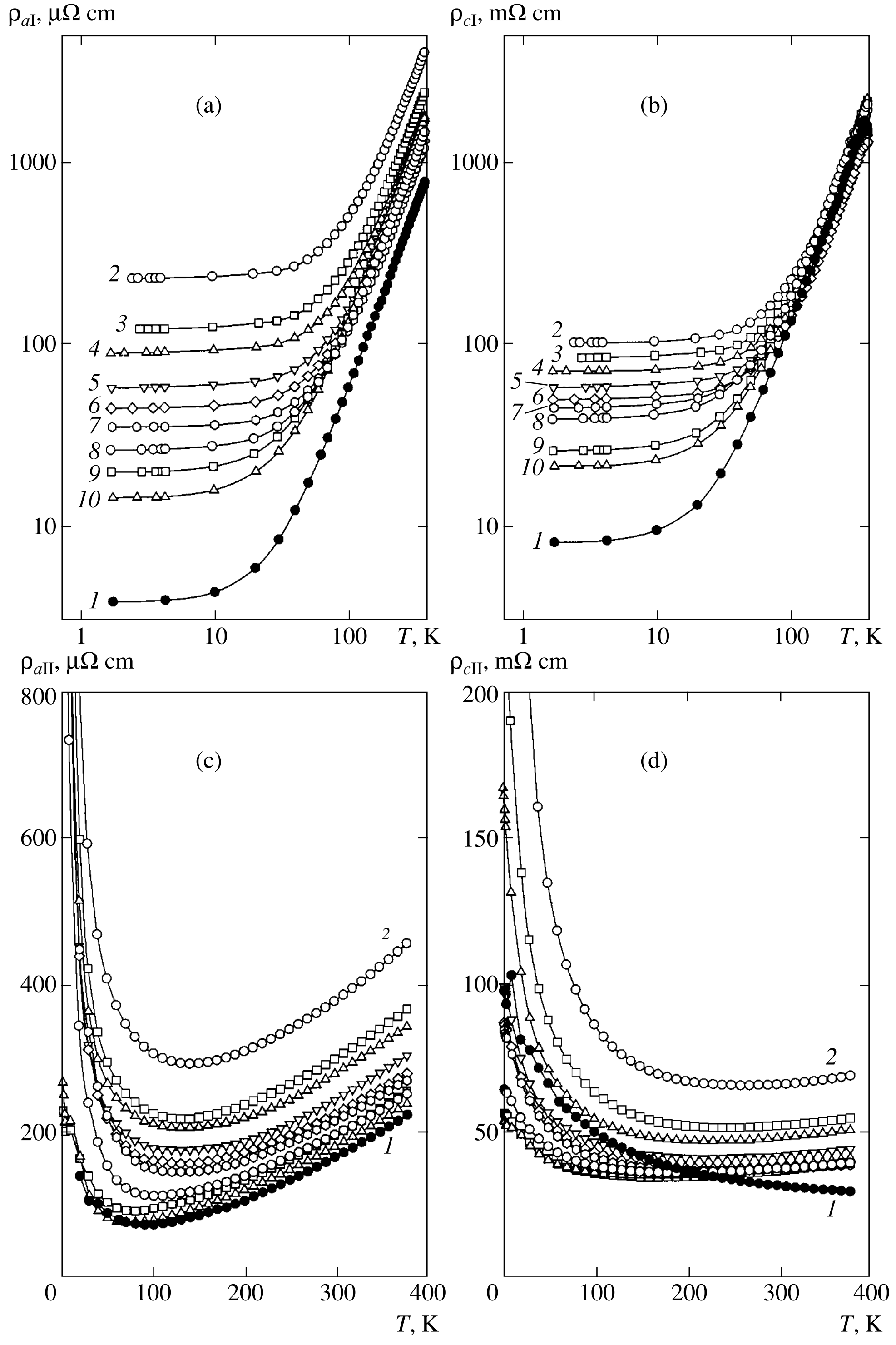


Fig. 6. The temperature dependences of the resistivity components (a) $\rho_{aI}(T)$, (b) $\rho_{cI}(T)$, (c) $\rho_{aII}(T)$, and (d) $\rho_{cII}(T)$ for an Sr_2RuO_4 crystal (*1*) in the initial (unirradiated) state, (*2*) upon irradiation with fast neutrons to a fluence of $\Phi = 3 \times 10^{19}$ cm^{-2}, and (*3*–*10*) after subsequent annealing at 300, 400, 500, 600, 700, 800, 900, and 1000°C, respectively. Curves connect the points corresponding to the same treatment.

charge carriers with relatively small effective mass m_a in accordance with the empirical relation $A^{\mathrm{II}}_{a2} \propto m_a^2$ [33]. However, the presence of the linear term $A^{\mathrm{II}}_{i1} T$ can be interpreted differently. Of course, this term could be related to the electron–phonon scattering (which is linear at high temperatures), but this type of scattering was rarely observed in systems with strong electron correlations. In HTSC-like systems, the linear variation of $\rho_a(T)$ observed in a broad temperature range (e.g., from

30 to 1300 K for $(La–Sr)_2CuO_4$) is interpreted as “non-Fermi” behavior of elementary excitations with allowance for the fact that the temperature dependence of the electric resistance is closer to linear than to quadratic (expected for the Fermi liquids). Similar behavior was also observed in many other systems with heavy fermions, for example, in $CeNi_2Ge_2$ [34] and $Sr_3Ru_2O_7$ (a double-layer analog of Sr_2RuO_4) [35].

Even more ambiguous is the physical meaning of the parameters A_{c1}^{II} and A_{c2}^{II}. Although the effective mass along the c axis may, in principle, differ for various parts of the Fermi surface determining the transport properties for the first and second conductivity channels, the nonsystematic behavior of A_{c2}^{II} upon the neutron irradiation and subsequent annealing (irreversibility) casts some doubt upon the possibility that the term $A_{c2}^{\mathrm{II}}T^2$ describes the electron–electron scattering. It should be noted that a positive slope of $\rho_c(T)$ (or $d\rho_c/dT > 0$) at high temperatures was also observed in less perfect crystals grown by zone melting under nonoptimum conditions or from a solution melt [36]. Therefore, it is possible that the irreversibility of changes in $\rho_c(T)$ observed in annealed samples is either related to some large-scale (nonpoint) defects formed in the course of high-temperature annealing as a result of the recombination of radiation defects, or this reversibility is caused by oxygen diffusion at temperatures above 800°C. Although the samples were annealed in air in a regime identical to the conditions of synthesis, the parameters of cooling and, hence, the crystal stoichiometry could hardly be the same for the synthesized and annealed samples.

The main result of the above analysis is that a channel of conductivity in which $d\rho/dT < 0$ not only for the charge transport along the c axis, but for the in-plane charge transfer as well, is present already in the initial (ordered) crystals of Sr_2RuO_4 rather than appears only upon irradiation or doping [29]. Thus, the compound studied features no electronic phase transitions: radiation-induced disorder on the atomic scale (as well as doping) accompanied by an increase in the residual resistivity (A_{i0}^{I}) only leads to a gradual decrease in the relative contribution of the first channel to the total conductivity. As a result, a region of negative slope in $\rho(T)$ (or $d\rho/dT < 0$) unavoidably appears. It should be noted that only the first channel may contribute, for example, to the de Haas–Van Alphen effect in the initial sample. In contrast, the second channel characterized by strong magnetic scattering (and, accordingly, small mean free path of charge carriers) at low temperatures cannot be studied using experimental methods employing the phenomenon of quantum oscillations.

Using the proposed model based on the concept of two conductivity channels with sharply different temperature dependences of the resistivity, it is possible to describe the behavior of $\rho_a(T)$ and $\rho_c(T)$ for the initial ordered sample at elevated temperatures (T = 400–1300 K). In particular, calculations using formulas (3) at T = 1000 K yield the resistivity components $\rho_a \approx$ 850 μΩ cm and $\rho_c \approx$ 30 mΩ cm, which agree well with the experimentally observed values of about 900 μΩ cm and 27 mΩ cm, respectively [26]. It should be noted that relations (3) offer only one variant of the description of experimental curves (Figs. 1 and 2) in terms of the two conductivity channels (bands). The expression for $\rho_{i\mathrm{II}}(T)$ appears formally as an expansion in powers of T (for $T_{i0} \sim 0$). However, the terms dominating at low and high temperatures are $B_{im}/(T + T_{i0})$ and $A_{i2}^{\mathrm{II}}T^2$, respectively, while the contributions of two other terms are not as significant. Apparently, formulas (3) offer the best variant for the description of experimental data within the framework of the model with two conductivity channels for a reasonable number of fitting parameters.

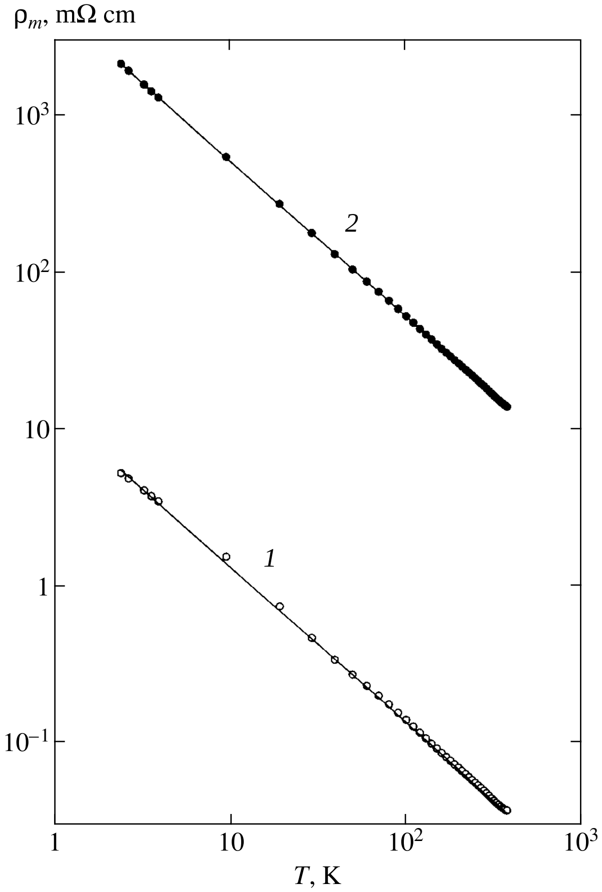


Fig. 7. The temperature dependence of magnetic contributions ρ_m via the second conductivity channel for the probing current directions (*1*) in plane and (*2*) along the c axis for an Sr_2RuO_4 crystal irradiated to a fluence of 3×10^{19} cm^{-2}. Curves connect the points corresponding to the same treatment.

We showed that radiation-induced disorder in Sr_2RuO_4 single crystals under fast neutron bombardment leads to the appearance of negative temperature coefficients ($d\rho_a/dT < 0$, $d\rho_c/dT < 0$) and negative magnetoresistances ($\Delta\rho_a < 0$, $\Delta\rho_c < 0$) at low temperatures. The magnetoresistance weakly depends on the probing current direction ($d\rho_a/dH \approx d\rho_c/dH$), but is strongly anisotropic with respect to the magnetic field orientation ($\mathbf{H} \parallel \mathbf{a}$ against $\mathbf{H} \parallel \mathbf{c}$).

The experimental curves of $\rho_a(T)$ and $\rho_c(T)$ for both the initial (ordered) samples and those radiation-disordered to various degrees of disorder can be described within the framework of a theoretical model with two channels of conductivity, which are characterized by strongly different temperature dependences of the resistivity. The first channel, in which $d\rho/dT < 0$ not only for the charge transport along the *c* axis but for the in-plane charge transfer as well, is present already in the initially ordered Sr_2RuO_4 crystals. The main contribution to the first channel is due to electrons (γ sheet of the Fermi surface) featuring electron–electron and impurity scattering, which leads to quadratic $\rho_a(T)$ and $\rho_c(T)$ functions. The second conductivity channel corresponds to the carriers possessing lower effective masses and featuring predominantly magnetic scattering at low temperatures, which account for the temperature dependence of the $\rho_{im}(T) \sim T^{-1}$ type.

ACKNOWLEDGMENTS

This study was supported in part by the Ministry of Industry and Technology of the Russian Federation (project nos. 40.012.1.1.1150 and 40.012.1.1.1356, contract no. 12/04), the Presidential Program of Support for Leading Scientific Schools in Russia (project no. NSh-639.2003.2), the Russian Foundation for Basic Research (project no. 04-02-16053), and the Program for Basic Research of the Presidium of the Russian Academy of Sciences (project no. 1000-251/P-03/040-348-11054-269) according to Grant No. 3 of the Ural Division of the Russian Academy of Sciences.

REFERENCES

1. S. L. Cooper and K. E. Gray, in *Physical Properties of High Temperature Superconductors IV,* Ed. by D. M. Ginsberg (World Sci., Singapore, 1994).
2. Y. Ando, G. S. Boebinger, A. Passner, *et al.*, Phys. Rev. Lett. **77**, 2065 (1996).
3. O. Laborde, M. Potel, P. Gougeon, *et al.*, Phys. Lett. A **147**, 525 (1990).
4. T. Kimura, S. Miyasaka, H. Takagi, *et al.*, Phys. Rev. B **53**, 8733 (1996).
5. Y. F. Yan, P. Matl, J. M. Harris, and N. P. Ong, Phys. Rev. B **52**, R751 (1995).
6. G. R. Stewart, Rev. Mod. Phys. **73**, 797 (2001).
7. F. Lichtenberg, A. Catana, J. Mannhart, and D. G. Schholm, Appl. Phys. Lett. **60**, 1138 (1992).
8. Y. Maeno, H. Hashimoto, K. Yoshida, *et al.*, Nature Lett. **372**, 532 (1994).
9. Y. Zha, S. L. Cooper, and D. Pines, Phys. Rev. B **53**, 8253 (1996).
10. B. A. Aleksashin, I. F. Berger, S. V. Verkhovskii, *et al.*, Physica C (Amsterdam) **153–155**, 339 (1988).
11. B. N. Goshchitskii, S. A. Davydov, A. E. Karkin, *et al.*, Physica C (Amsterdam) **162–164**, 1019 (1989).
12. B. A. Aleksashin, V. I. Voronin, S. V. Verkhovskiĭ, *et al.*, Zh. Éksp. Teor. Fiz. **95**, 678 (1989) [Sov. Phys. JETP **68**, 382 (1989)].
13. A. E. Kar'kin, V. V. Shchennikov, B. N. Goshchitskiĭ, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **45**, 2147 (2003) [Phys. Solid State **45**, 2249 (2003)].
14. A. P. Mackenzie, S. R. Julian, A. J. Diver, *et al.*, Phys. Rev. Lett. **76**, 3786 (1996).
15. C. Bergemann, S. R. Julian, A. P. Mackenzie, *et al.*, Phys. Rev. Lett. **84**, 2662 (2000).
16. S. Hill, J. S. Brooks, Z. Q. Mao, and Y. Maeno, Phys. Rev. Lett. **84**, 3374 (2000).
17. T. Yokoya, A. Chainani, T. Takahashi, *et al.*, Phys. Rev. Lett. **76**, 3009 (1996).
18. H. Yoshino, K. Murata, N. Shirakawa, *et al.*, J. Phys. Soc. Jpn. **65**, 1548 (1996).
19. M. A. Tanatar, M. Suzuki, S. Nagai, and T. Ishiguro, Phys. Rev. Lett. **86**, 2649 (2001).
20. V. G. Hildebrand, M. Reedyk, T. Katsufuji, and Y. Tokura, Phys. Rev. Lett. **87**, 227002 (2001).
21. T. Imai, A. W. Hunt, K. R. Thurber, and F. C. Chou, Phys. Rev. Lett. **81**, 3006 (1998).
22. T. Katsufuji, M. Kasai, and Y. Tokura, Phys. Rev. Lett. **76**, 126 (1996).
23. S. Sakita, S. Nimori, Z. Q. Mao, *et al.*, Phys. Rev. B **63**, 134520 (2001).
24. T. Oguchi, Phys. Rev. B **51**, 1385 (1995); D. J. Singh, Phys. Rev. B **52**, 1358 (1995).
25. A. P. Mackenzie, N. E. Hussey, A. J. Diver, *et al.*, Phys. Rev. B **54**, 7425 (1996).
26. A. W. Tyler, A. P. Mackenzie, S. Nishizaki, and Y. Maeno, Phys. Rev. B **58**, R10107 (1998).
27. N. E. Hussey, A. P. Mackenzie, J. R. Cooper, *et al.*, Phys. Rev. B **57**, 5505 (1998).
28. R. Jin, Y. Liu, and F. Lichtenberg, Phys. Rev. B **60**, 10418 (1999).
29. M. Minakata and Y. Maeno, Phys. Rev. B **63**, 180504 (2001).
30. H. C. Montgomery, J. Appl. Phys. **42**, 2971 (1971).
31. E. Miranda, V. Dobrosavljevic, and G. Kotliar, J. Phys.: Condens. Matter **8**, 9871 (1996).
32. B. Andraka, Phys. Rev. B **49**, 3589 (1994).
33. Y. Maeno, K. Yoshida, H. Hashimoto, *et al.*, J. Phys. Soc. Jpn. **66**, 1405 (1997).
34. S. R. Julian, C. Pfleiderer, F. M. Groshe, *et al.*, J. Phys.: Condens. Matter **8**, 9675 (1996).
35. Y. Liu, R. Jin, Q. Mao, *et al.*, Phys. Rev. B **63**, 174435 (2001).
36. H. Berger, L. Forro, and D. Pavuna, Europhys. Lett. **41**, 531 (1998).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 100, No. 6, 2005, pp. 1153–1158.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 6, 2005, pp. 1310–1316.
Original Russian Text Copyright © 2005 by Ritus, Volkov.

ELECTRONIC PROPERTIES OF SOLIDS

Photoinduced Submillimeter-Wave Amplitude Diffraction Grating in a Semiconducting CdF_2:Ga Crystal

A. I. Ritus and A. A. Volkov

Institute of General Physics, Russian Academy of Sciences, ul. Vavilova 38, Moscow, 119991 Russia

e-mail: ritus@ran.gpi.ru

Received February 4, 2005

Abstract—It is shown that exposure of an additively colored CdF_2:Ga crystal with bistable *DX* centers that is slowly cooled to 150 K to blue-green light through a slotted mask produces a submillimeter-wave diffraction grating, which persists for a long time at temperatures of 160–240 K. It is also shown that the diffraction grating induced in a sample is an amplitude grating. The absorption of submillimeter waves in illuminated regions of the sample is associated with the conductivity due to the transition of impurity centers to a metastable donor state. In the *n*–*i*–*n*–*i*-type conducting structure obtained, the conductivity of *n*-type regions at 225 K amounts to $\sigma' \approx 0.24\ \Omega^{-1}\ \text{cm}^{-1}$. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

A CdF_2 crystal is an ionic dielectric with a wide band gap of about 7.8 eV; it converts into an *n*-type semiconductor under doping by trivalent impurities and subsequent annealing in alkaline metal vapor [1]. In Ga doped CdF_2 subjected to such a procedure of additive coloring, impurity centers may be in two states, which are characteristic of *DX* centers. The shallow state corresponds to the localization of an electron on a hydrogen-like orbit ($Ga^{3+} + e$). In the ground, deep state, two electrons are localized on the impurity ion (Ga^{1+}), and the ion itself is displaced to the adjacent interstitial site, thus giving rise to a local distortion of the lattice. The corresponding potential barrier between shallow and deep states is responsible for the metastability of the shallow state and for the difference between the thermal and optical ionization energies of the deep state. The depth of the shallow impurity level is about 0.116 eV [2], and that of the deep level is about 0.75 eV [3]; the optical ionization energy of the deep center is about 3 eV [4].

At a temperature below $T \approx 240$ K [4, 5], metastable shallow centers that are induced by the light quanta $h\nu \approx 3$ eV during the reaction

$$Ga^{1+} + Ga^{3+} + h\nu \longrightarrow 2(Ga^{3+} + e) \qquad (1)$$

may persist for a long time. This transition is accompanied by a decrease in the refractive index Δn in the frequency range between the visible ultraviolet absorption band of the deep center ($\lambda_{max} \approx 400$ nm) and the infrared band of the shallow center ($\lambda_{max} \approx 7\ \mu$m). Owing to the locality and the metastability of variations in Δn and their dependence on the intensity of light, one can use a CdF_2:Ga crystal for recording optical holograms [6]. As temperature increases, the photoinduced shallow centers decay according to the reverse reaction

$$2(Ga^{3+} + e) + kT \longrightarrow Ga^{1+} + Ga^{3+}, \qquad (2)$$

and the photoinduced variations in Δn vanish.

As temperature decreases, the impurity ions gradually pass to the deep state. However, when a CdF_2:Ga crystal is cooled sufficiently rapidly to temperatures below 240 K, part of the impurity centers remain in the shallow state due to the potential barrier between the shallow and deep states.

The temperature and illumination dependence of the refractive index similar to that described above for CdF_2:Ga is also characteristic of other semiconductors with bistable *DX* centers [7–10] that were used earlier as the recording media for phase holograms in the frequency range between the absorption bands of the deep and shallow centers. Usually, interference fringes are recorded at the intersection of two laser beams, and these holograms represent a phase diffraction grating. In this paper, we show that, in an additively colored CdF_2:Ga crystal with bistable *DX* centers, one can obtain a photoinduced amplitude diffraction grating in the frequency range below the absorption band of shallow centers.

Such a possibility arises in a temperature interval low enough that the decay of photoinduced shallow centers according to reaction (2) does not occur and, at the same time, high enough that the thermal ionization of shallow centers creates a significant concentration of free electrons. In the case of a CdF_2:Ga crystal, this condition is fulfilled at temperatures of 150–240 K. The local character of the conductive structure produced and its correspondence to the profile of the illumination

1063-7761/05/10006-1153$26.00 © 2005 Pleiades Publishing, Inc.

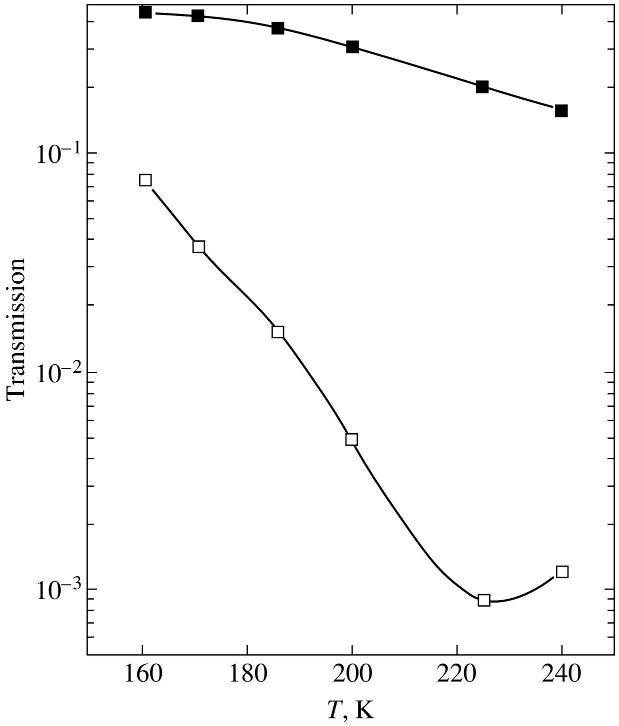


Fig. 1. Temperature dependence of the submillimeter-wave transmission of a CdF_2:Ga sample before (filled squares) and after (open squares) exposure to blue-green light (l = 2.18 mm and λ = 0.65 mm).

intensity are guaranteed by the small value of electron mobility, $\mu \approx 15$ cm^2 V^{-1} s^{-1} in CdF_2 : Ga [11] at these temperatures, and the Coulomb attraction of an ionized *DX* center. For these values of the mobility, the mean time between electron scattering events (with the effective mass $m^* = 0.45m_e$ for CdF_2 [11]) is

$$\tau = \mu m^*/e = 4\times 10^{-15}\ \mathrm{s},$$

and the corresponding mean free path is

$$l = v_{\mathrm{therm}}\tau = (3kT/m^*)^{1/2}\tau$$
$$\approx 5\times 10^{-8}\ \mathrm{cm},$$

which is much less than the wavelength (4×10^{-5} cm) of the illuminating light. In this case, the real part of the Drude conductivity is

$$\sigma' = \frac{\sigma_0}{1+\omega^2\tau^2}, \quad \sigma_0 = N_e e\mu, \tag{3}$$

and the corresponding absorption coefficient is

$$\alpha = \frac{\sigma'}{\varepsilon_0 cn} \tag{4}$$

and remains practically unchanged up to a frequency of

$$\nu_D = \frac{1}{2\pi\tau} = 4\times 10^{13}\ \mathrm{Hz},$$

i.e., up to the mid infrared band (here, N_e is the concentration of conduction electrons, e is the electron charge, ε_0 is the permittivity of vacuum, c is the velocity of light, and n is the refractive index). As we showed in [12], when a CdF_2:Ga crystal is cooled sufficiently slowly to T = 200 K, the concentrations of ions in the shallow state N_{sh} before and after photoexposition may differ by a factor greater than 20. Therefore, the concentrations of conduction electrons N_e and the absorption coefficients α in the illuminated and unilluminated parts of the sample may also differ by a factor of more than 20. Thus, a photoinduced amplitude diffraction grating in CdF_2:Ga must have high contrast and be reliably recorded in the submillimeter-wave band.

2. DESCRIPTION OF THE EXPERIMENT

Experiments were carried out on the submillimeter-wave spectrometer Epsilon [13]. First, we measured the transmission coefficient *Tr* versus temperature of a plane-parallel sample of CdF_2:Ga of thickness l = 2.18 mm at a wavelength of λ = 0.65 mm in the dark and after exposing it to blue-green light. To this end, we put caps with windows of black polyethylene film on the windows of a cryostat; this film well transmits submillimeter waves and cuts off visible light. Under one of the caps, we mounted two light-emitting diodes (LEDs) above and below the aperture of the polyethylene window to illuminate the sample by blue-green light.

The temperature dependence of the transmission of CdF_2:Ga in the dark at a wavelength of λ = 0.65 mm was recorded while the sample was cooled from room temperature to a temperature of 150 K. First, from 300 to 263 K, the sample was cooled rapidly. In the interval from 263 to 225 K, which included the metastability temperature $T \approx 240$ K, the sample was cooled slowly at a rate of 0.3 K/min, and then, in the range of temperatures 225–150 K, the sample was cooled at a rate of 0.7 K/min.

At a temperature of T = 150 K, the sample was illuminated for 25 min by blue-green light emitted by two LEDs; after that, the transmission of the sample with photoinduced delayed conductivity was measured in the dark at a wavelength of λ = 0.65 mm in the temperature interval from 150 to 240 K. These measurements were carried out while the sample was heated during 1 h and 15 min.

The measured submillimeter-wave transmission of CdF_2:Ga is shown in Fig. 1. The difference between the transmissions of the illuminated and unilluminated samples of CdF_2:Ga attains its maximum at a temperature of about 225 K. Using the formulas

$$Tr = (1-R)^2\exp(-\alpha l), \quad R = \frac{(n-1)^2}{(n+1)^2} \tag{5}$$

for the transmission of a sample and formula (4), we find that, for $n = \sqrt{\varepsilon'} \approx 3$ [14, 15] and Tr (225 K) = 0.0009 for an illuminated sample, the delayed conductivity equals σ' (225 K) = 0.24 Ω^{-1} cm^{-1}.

To carry out experiments with a diffraction grating photoinduced in a CdF_2:Ga sample, we modified the Epsilon spectrometer to allow for the measurement of the angular dependence of the intensity of submillimeter waves transmitted through a sample.

A quasi-parallel beam of radiation with a wavelength of $\lambda = 0.65$ mm emitted by a backward wave oscillator was focused on a sample by a teflon lens with a focal length of $F = 340$ mm; the half-width of the Gaussian intensity profile on the sample was about 9 mm. A metal diaphragm with a 8 × 8-mm square hole was placed behind the sample so that the intensity of the submillimeter-wave radiation on the sample was approximately constant within the aperture of the diaphragm. A transparent film was sandwiched between the sample and the diaphragm. On this film (from the side of the sample), black parallel vertical strips 2 mm in width were printed with a step of 4 mm by a laser printer, so that two black and two transparent strips fall within the 8 × 8-mm aperture. The whole structure was placed in a cryostat with caps of black polyethylene put on its windows. The cap with LEDs was placed on the side where the film with black strips was attached. When the LEDs were switched on, the regions of the sample under transparent strips were illuminated, while the regions under black strips were practically unilluminated. The sharpness of the shadows from the strips was guaranteed by the configuration of the LEDs, which were arranged along a vertical axis at a distance of about 7 cm from the sample. We carried out a separate experiment to find out that, at $\lambda = 0.65$ mm, the transmission of the film regions under the black strips is the same as that outside these strips; i.e., the black strips on the film are practically transparent for submillimeter waves.

To measure the angular dependence of the intensity of submillimeter waves transmitted through the sample, we placed the receiver (a Golay cell) of the spectrometer at the end of a rod that could rotate about a vertical axis passing through the sample. A lens with the focal length $F = 120$ mm was placed at the middle of the rod; the distance between the lens and either the sample or the receiver was about twice the focal length. Such a configuration guaranteed a sufficiently high degree of sensitivity and the necessary angular resolution. The aperture of the cryostat windows was large enough to allow for the angular measurements within ±16°. For the chosen configuration of two 2-mm-wide slots with a step of 4 mm on an opaque screen, the first-order diffraction maxima for $\lambda = 0.65$ mm are located at about ±8.5° and amount to 44% of the intensity of the zeroth-order maximum, and the second-order diffraction maxima amount to only 1% of the intensity of the zeroth-order maximum. Thus, practically the whole diffracted radiation falls within the aperture of the cryostat window.

To carry out experiments with the photoinduced diffraction grating, we used the same 2.18-mm-thick $CdF_2:Ga$ sample and followed the same cooling regime as in the preliminary measurements of transmission versus temperature in the dark. At a temperature of 150 K, the sample was illuminated for 25 min by blue-green light from two LEDs. Then, we measured the angular distribution of the intensity of submillimiter waves ($\lambda = 0.65$ mm) transmitted through the sample at temperatures of 161, 171, 186, 200, 225, and 240 K in the dark. These measurements were carried out during heating of the sample with necessary stops at each temperature point and took 2.5 hours.

3. EXPERIMENTAL RESULTS AND DISCUSSION

Figure 2 shows the angular distribution of diffracted radiation measured at different temperatures. Each distribution is normalized by the intensity of the zeroth-order maximum at the relevant temperature. For the sake of clarity, each diagram is shifted by 0.3 along the vertical axis. The symbols represent the experimental results with a step of 1°, and the solid curves represent the theoretical results (see below) calculated for the transmission in the dark and after exposure (see Fig. 1).

The lowest diagram is obtained at 150 K before exposing the sample to blue-green light; i.e., this distribution corresponds to the diffraction of radiation with $\lambda = 0.65$ mm by a single 8-mm-wide slot in an opaque screen. For such a configuration, the angular distribution of radiation intensity $I(\theta)$ behind the slot is given by [16]

$$I(\theta) = \sin^2 u/u^2, \quad u = \pi a \sin\theta/\lambda, \tag{6}$$

where a is the slot width and θ is the angle between the direction of the incident beam and the direction to the receiver. The solid (theoretical) curve for this case coincides with the experimental data, which means, in particular, that the receiving system has a sufficiently high angular resolution.

The angular distributions of diffracted radiation obtained after exposing the sample to blue-green light through the mask exhibit pronounced peaks near ±8.5° at all temperature points, which corresponds to the first-order diffraction maxima for the grating consisting of two 2-mm-wide slots with a step of 4 mm in an opaque screen. The relative intensity of these peaks increases

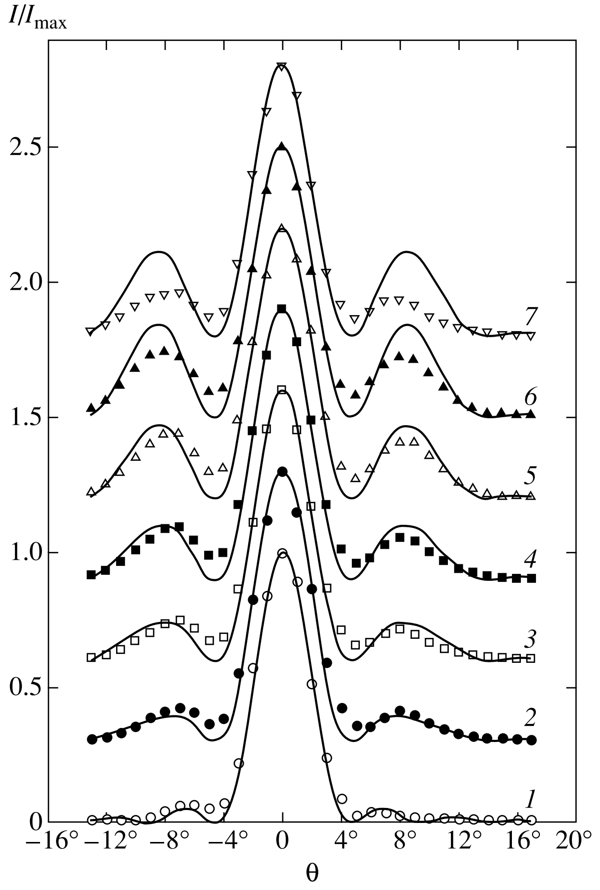


Fig. 2. Angular distributions of submillimeter-wave radiation diffracted by a photoinduced grating in CdF_2:Ga; (*1*) before exposure, (*2–7*) after exposure at different temperatures: T = (*2*) 161, (*3*) 171, (4) 186), (5) 200, (*6*) 225, and (*7*) 240 K. Every subsequent graph is shifted by 0.3 along the vertical axis. The symbols represent experimental results, and the solid lines represent the results of calculation by formulas (6)–(11) for the values of transmission in the dark and after exposure, which are shown in Fig. 1 (l = 2.18 mm, a = 2 mm, d = 4 mm, N = 2, and λ = 0.65 mm).

with temperature up to T = 225 K and decreases at T = 240 K. Qualitatively, this result correlates with the behavior of the contrast of the transmissions of the illuminated and unilluminated regions of a CdF_2:Ga sample (see Fig. 1). At T = 225 K, the relative intensity of the first-order maxima amounts to 23%, which is about one half the intensity of these maxima (44%) in the case of two 2-mm-wide slots with a step of 4 mm in an opaque screen. Obviously, in the case of our structure, which consists of alternating strips with transmissions Tr_1 and Tr_2, where the indices 1 and 2 refer to the unilluminated and illuminated regions, respectively, the relative intensity of the first-order maxima must be lower due to the additional contribution of the strips with transmission $Tr_2 \neq 0$ to the diffraction pattern.

We calculated the diffraction pattern for a periodic structure consisting of alternating strips with transmissions Tr_1 and Tr_2 in the case when the number of strips of the fist type, N_1, is equal to the number of strips of the second type, N_2, i.e., $N_1 = N_2 = N$, in the following manner. The diffraction by such a structure can be considered as interference between the diffraction fields from two gratings: a grating of N slots with transmission Tr_1 in an opaque screen and a grating of N slots with transmission Tr_2 in an opaque screen, the latter grating being shifted by half the period d with respect to the first one.

According to the Fresnel–Kirchhoff diffraction theory [16], the amplitude E_1 of the field from the first grating is determined by the expressions

$$\begin{gathered} E_1 = b\frac{t_1 a_1 \sin u_1 \sin(N\nu)}{u_1 \sin\nu}, \quad t_1 = \sqrt{Tr_1}, \\ u_1 = \pi a_1 \sin\theta/\lambda, \quad \nu = \pi d \sin\theta/\lambda, \end{gathered} \tag{7}$$

where b is a dimensional proportionality factor and a_1 is the width of the slots of the first grating. Similarly, the amplitude E_2 of the field from the second grating is determined by the expressions

$$\begin{gathered} E_2 = b\frac{t_2 a_2 \sin u_2 \sin(N\nu)}{u_2 \sin\nu}, \quad t_2 = \sqrt{Tr_2}, \\ u_2 = \pi a_2 \sin\theta/\lambda, \end{gathered} \tag{8}$$

where a_2 is the width of the slots in the second grating. The oscillation phases of the fields E_1 and E_2 differ by

$$\varphi = \frac{2\pi d}{\lambda\ 2}\sin\theta. \tag{9}$$

The amplitude of the field $E(\theta)$ of the resulting diffraction pattern with regard to interference is given by

$$E^2(\theta) = E_1^2 + E_2^2 + 2E_1E_2\cos\varphi; \tag{10}$$

accordingly, the intensity $I(\theta)$ in the direction θ is given by

$$\frac{I(\theta)}{I(0)} = \frac{E^2(\theta)}{E^2(0)}. \tag{11}$$

Thus, the coefficient b cancels out in the expression for the relative intensity.

The results of calculations by formulas (7)–(11) using the experimental parameters $a_1 = a_2$ = 2 mm, d = 4 mm, N = 2, and λ = 0.65 mm and the measured values Tr_1 and Tr_2 of transmission in the dark and after exposure are shown in Fig. 2. One can see that the experimental results are well described by computed curves at temperatures 161–200 K, whereas, at temperatures T = 225 K and T = 240 K, the experimental points for the first-order maxima lie significantly lower than the computed results. Figure 3 shows the temperature dependence of the relative intensity of the first-order maxima obtained in the experiment (filled squares) and calculated by formulas (7)–(11) (open squares). One can see that, despite a certain quantitative discrepancy at temperatures of 220–240 K, the experimental and theoretical functions are qualitatively identical: first, the relative intensity of the first-order maxima increases with temperature due to the increased contrast of the grating because of the increased density of electrons in the illuminated strips of the sample, and then the diffraction maxima decrease due to the decreased population of the metastable state and the associated drop in the electron density. Note that the maxima of both curves occur near the same temperature point of $T \approx 220$ K. This correlation between experimental results and the theoretical calculations based on the transmission of illuminated and unilluminated strips of the sample provide reliable evidence for the presence of a photoinduced amplitude diffraction grating in the sample.

In our view, the main reason for the quantitative discrepancy, especially in the range from 220 to 240 K, is the decay of the metastable state at temperatures close to 240 K. For instance, according to [4], at $T = 254$ K, the infrared absorption from the metastable level in CdF_2:Ga is halved in 600 s after exposure. Our measurements show that, at $T = 240$ K, the submillimeter-wave conductivity is halved in 1000 s after exposure. Hence, at temperatures slightly below 240 K, the half-life time may be on the order of 1 h. Our preliminary measurements of the transmission of the CdF_2:Ga sample during heating after exposure took 1 h and 15 min, whereas the measurements of the angular distributions of the diffraction pattern took twice as much time. Thus, in the diffraction experiment, when approaching temperatures of 225 and 240 K, a metastable state was depleted over a longer period, the transmission Tr_2 was greater, and the contrast of the grating was weaker compared with the preliminary measurements.

The reduced contrast of exposure due to the incomplete absorption of light by black strips of the mask, scattering of light in the sample, and the fact that LEDs are not point sources may constitute the second reason for the above quantitative discrepancy.

In [17], the authors observed significant photoinduced variations in the dielectric response of CdF_2:Ga and CdF_2:In crystals at a wavelength of $\lambda = 8$ mm. They attribute these variations to the resonance absorption in the infrared band due to the ionized donor pairs. The variation in ε' in CdF_2:Ga at $T = 1.8$ K was $\Delta\varepsilon' = 0.5 \pm 0.1$. If we assume that there is a similar photoinduced variation in ε' at $\lambda = 0.65$ mm at temperatures of 150–200 K, then this should give rise to an additional phase shift $\Delta\varphi$ between submillimeter-wave beams transmitted through the illuminated and unilluminated strips of a sample of thickness l:

$$\Delta\varphi = \frac{2\pi}{\lambda} l(\sqrt{\varepsilon_1'} - \sqrt{\varepsilon_2'}), \qquad (12)$$

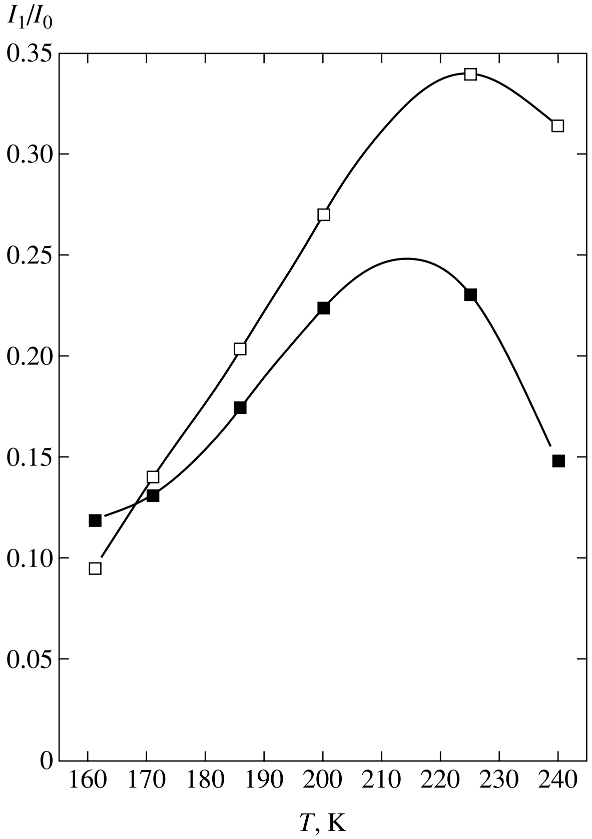


Fig. 3. Temperature dependence of the intensity of the diffraction maxima of orders –1 and +1. Filled and open squares represent, respectively, experimental results and the results of calculations by formulas (7)–(11) for the values of transmission in the dark and after exposure shown in Fig. 1.

where the indices "1" and "2" refer to the illuminated and unilluminated areas and $\varepsilon_2' = \varepsilon_1' + \Delta\varepsilon'$. When calculating diffraction by a photoinduced grating, one should add this phase shift to the argument of cosine in (10). We performed such a calculation for the sample temperature 161 K at which the transmission contrast between the illuminated and unilluminated areas is minimal (see Fig. 1) and, accordingly, the effect of the additional phase shift is maximal. The results of calculations for $l = 2.18$ mm and ε_1' (161 K) ≈ 9 [14, 15] are presented in Fig. 4 for $\Delta\varphi = -0.55\pi$, which corresponds to $\Delta\varepsilon' = 0.5$, and for $\Delta\varphi = -0.056\pi$, which corresponds to $\Delta\varepsilon' = 0.05$. One can see that the calculated curves are characterized by asymmetric distribution of the intensity and the shape of the diffraction maxima of orders +1 and –1, whereas the experimental curves are symmetric.[1] For $\Delta\varepsilon' = 0.5$, the calculations also show a multifold increase in the amplitudes of the maxima of orders –1 and +1 and a shift in the zeroth-order maxi-

---

[1] A small difference between the intensities of the maxima of orders –1 and +1 is attributed to the apparatus effect, which was observed even before illuminating the sample (see Fig. 2).

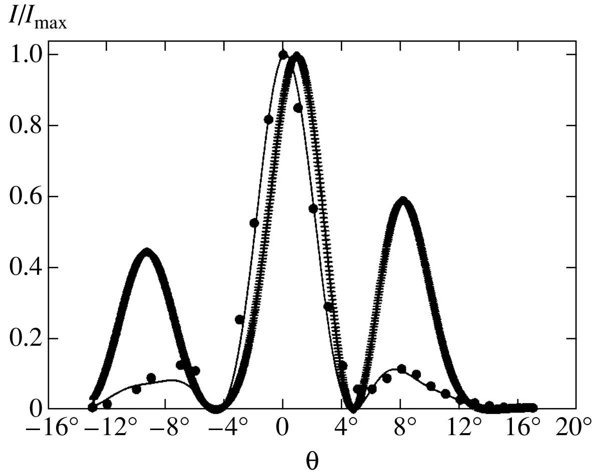


Fig. 4. Diffraction patterns of a photoinduced grating in CdF_2:Ga at $T = 161$ K. Filled circles represent the experimental distributions of intensity, and the crosses and the solid curve represent the results of calculations performed with regard to the phase shift (formula (12)) for $\Delta\varepsilon' = 0.5$ and 0.05, respectively.

mum by about 1° compared with the experimental data. Thus, our experiment shows that, at $\lambda = 0.65$ mm and temperatures of 150–240 K, a photoinduced variation in ε' is extremely small ($\Delta\varepsilon' < 0.05$). However, note that the angular distributions of the diffraction pattern for $\Delta\varepsilon' \approx 0$ coincide with the distributions for which the phase difference is $\Delta\varphi \approx -2\pi k$, $k = 1, 2, 3, \ldots$ (see formula (10)). If some of these low-probability situations happen to occur, then the corresponding $\Delta\varepsilon' \approx 1.9$, or 3.9, or 6.2, etc.

4. CONCLUSIONS

Thus, we have shown that the exposure of a CdF_2:Ga crystal, slowly cooled in the dark to 150 K, to blue-green light through a slotted mask produces a submillimeter-wave diffraction grating, which persists for a long time at temperatures of 160–240 K. The agreement between the experimental temperature dependence of the intensities of diffraction maxima and theoretical calculations of the diffraction pattern based on the temperature dependence of the transmission of illuminated and unilluminated strips in a sample provide strong evidence that the diffraction grating induced by light in the sample is an amplitude grating.

The absorption of submillimeter waves in the illuminated regions of the sample is attributed to the Drude conductivity of electrons thermally activated from a shallow level of impurity centers transferred to a metastable donor state. Actually, we have obtained a photoinduced long-term conducting *n*–*i*–*n*–*i*-type structure; at 225 K, the *n*-region conductivity reaches a value of $\sigma' \approx 0.24\ \Omega^{-1}\ \text{cm}^{-1}$. It is obvious that, using different masks, one can obtain photoinduced long-term structures of arbitrary shape in a CdF_2:Ga sample (and in other materials with bistable *DX* centers), the width of conducting strips being of submicron range.

At $\lambda = 0.65$ mm, there is no appreciable photoinduced variation in the dielectric constant ε' in CdF_2:Ga at temperatures of 150–240 K ($\Delta\varepsilon' < 0.05$) because we did not observe the corresponding characteristic variations in the diffraction pattern in our experiments.

ACKNOWLEDGMENTS

We are grateful to G.A. Komandin for useful recommendations and to colleagues S.A. Kazanskii, A.I. Ryskin, and A.S. Shcheulin from the State Optical Institute for cooperation.

This work was supported by the International Science and Technology Center, project no. 2136.

REFERENCES

1. J. D. Kingsley and J. S. Prener, Phys. Rev. Lett. **8**, 315 (1962).
2. J. M. Langer, T. Langer, G. L. Pearson, *et al.*, Phys. Status Solidi B **66**, 537 (1974).
3. C. H. Park and D. J. Chadi, Phys. Rev. Lett. **82**, 113 (1999).
4. A. I. Ryskin, A. S. Shcheulin, and D. E. Onopko, Phys. Rev. Lett. **80**, 2949 (1998).
5. B. Koziarska-Glinka, A. Barcz, L. Arizmendi, and A. Suchocki, Phys. Rev. B **61**, 9295 (2000).
6. A. Suchocki, B. Koziarska, T. Langer, and J. M. Langer, Appl. Phys. Lett. **70**, 2934 (1997).
7. R. A. Linke, T. Thio, J. D. Chadi, and G. E. Devlin, Appl. Phys. Lett. **65**, 16 (1994).
8. R. A. Linke, T. Thio, J. D. Chadi, *et al.*, in *Proceedings of the 22nd International Conference on the Physics of Semiconductors* (World Sci., London, 1995).
9. T. Thio, J. W. Bennett, D. J. Chadi, *et al.*, J. Cryst. Growth **159**, 345 (1996).
10. A. I. Ryskin, A. S. Shcheulin, B. Koziarska, *et al.*, Appl. Phys. Lett. **67**, 31 (1995).
11. R. P. Khosla and D. Matz, Solid State Commun. **6**, 859 (1968); R. P. Khosla, Phys. Rev. **183**, 695 (1969).
12. V. V. Kasparov, A. A. Volkov, and A. I. Ritus, Izv. Ross. Akad. Nauk, Ser. Fiz. **67**, 1763 (2003).
13. A. A. Volkov, Yu. G. Goncharov, G. V. Kozlov, *et al.*, Infrared Phys. **25**, 369 (1985); G. V. Kozlov and A. A. Volkov, in *Topics in Applied Physics*, Ed. by G. Gruner (Springer, Berlin, 1998), Vol. 74, p. 51.
14. A. I. Ritus, A. V. Pronin, A. A. Volkov, *et al.*, Phys. Rev. B **65**, 165209 (2002).
15. J. D. Axe, J. W. Gaglianello, and J. E. Scardefield, Phys. Rev. **139**, A1211 (1965).
16. M. Born and E. Wolf, *Principles of Optics*, 4th ed. (Pergamon, Oxford, 1969; Nauka, Moscow, 1973), Chap. 8.
17. S. A. Kazanskii, D. S. Rumyantsev, and A. I. Ryskin, Phys. Rev. B **65**, 165214 (2002).

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 100, No. 6, 2005, pp. 1159–1174.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 6, 2005, pp. 1317–1333.
Original Russian Text Copyright © 2005 by Brantov, Bychenkov, Rozmus, Capjack.

Permittivity of Plasma and Nonstationary Theory of Nonlocal Transport

A. V. Brantov[a], V. Yu. Bychenkov[a], W. Rozmus[b], and C. E. Capjack[c]
[a]*Lebedev Physical Institute, Russian Academy of Sciences, Moscow, 119991 Russia*
[b]*Department of Physics, University of Alberta, Edmonton, Alberta, T6G 2J1, Canada*
[c]*Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Alberta, T6G 2J1, Canada*
e-mail: brantov@sci.lebedev.ru; abrantov@phys.ualberta.ca

Received November 11, 2004

Abstract—A regular procedure is proposed for finding the solution to a linearized kinetic equation for charged particles with the Landau collision integral in a plasma with large *Z*. The expression for longitudinal permittivity of a collisional plasma, which is obtained using this procedure for the entire range of frequencies and wavenumbers, as well as the collision parameter, is transformed to the known expressions in the corresponding asymptotic limits. The nonlocal transport equations for small perturbations are also formulated for arbitrary relations between the characteristic space and time scales of the plasma; these relations considerably extend the limits of applicability for previously developed theories. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

In spite of the fact that dielectric susceptibility of plasma is a parameter invariably considered in all textbooks on plasma physics, a universal expression or a simple algorithm for its derivation for collisional plasma has not been obtained for the entire range of wavenumbers k and frequencies ω. This is due to the fact that even the determination of a linear response of the plasma involves the solution of an integrodifferential kinetic equation for particles experiencing Coulomb collisions. The derivation of such a solution in a form permitting its routine use in various applications is the main constraint in the theory of plasma response with collisions. A natural simplification of the theory becomes possible when the exact Landau collision integral is replaced by model expressions or when simplifying assumptions concerning weak or, conversely, strong collisions of particles may lead to a noticeable loss in the accuracy of computations.

The most widely used expression for the permittivity of a collisional plasma can easily be derived by using the model Bhatnagar–Gross–Krook (BGK) collision integral [1]. This approximation makes it possible to qualitatively describe the effect of collisions on dispersion properties of the plasma. However, the use of permittivity for the BGK model in a certain region of (ω, k) leads to a significant error. Attempts to improve such a model description by introducing the Rutherford dependence of the effective collision frequency on the velocity have not substantially improved the accuracy in determining the dielectric susceptibility [2, 3]. A noticeable improvement of the theory was obtained when the electrostatic response of a plasma was determined using the Lorentz model with an exact (in parameter m_e/m_i) Landau electron–ion collision integral [4–6]. This was manifested in a narrowing of the (ω, k) region in which a satisfactory quantitative description of the dielectric constant cannot be obtained. At the same time, it will be shown below that disregard of the electron–electron collision integral still does not allow obtaining a quantitative description of the dielectric properties of plasmas in the entire range of frequencies and wavelengths.

In addition to theoretical models providing a unified quantitative description of the dielectric susceptibility in the entire range of (ω, k), standard approaches of perturbation theory give correct asymptotic concepts of dielectric susceptibility. Such approaches include the kinetic theory of the response of a weakly collisional plasma [7] and hydrodynamic-type theory for collisional plasma [8]. These theories naturally have strong limitations in parameters ω/ν_{ei} and $k\lambda_{ei}$, where ν_{ei} is the frequency of electron–ion (e–i) collisions and λ_{ei} is the mean free path of electrons. Although the weakly collisional nonlocal theory proposed in [9] expands the region of qualitative description of permittivity, it still fails to cover all space and time scales of perturbations. The most direct method for calculating the dielectric susceptibility for arbitrarily set values of (ω, k) involves numerical solution of the Fokker–Planck kinetic equation in the Fourier representation, which is quite nontrivial, requires practical skills in simulation and cannot be easily parametrized. Consequently, the construction of a theory providing a universal method for reproducing the dielectric susceptibility for the entire range of frequencies and wavenumbers (ν_{ei}, λ_{ei}) and the collision parameter of a plasma is a problem important for practical applications. The present study is devoted to the solution of this problem.

The problem of dielectric susceptibility of a collisional plasma is closely related to the problem of nonlocal transport. Theoretical models of nonlocal transport in hot plasmas have been being developed for more than 20 years beginning from publications [10–12]; however, the improvement of these models is still not over for the condition $\lambda_{ei}/L < 10^{-2}$, which is typical, on the one hand, for inertial confinement fusion (ICF) experiments and, on the other hand, in cases when the classical strongly collisional theory of transport is inapplicable [13, 14]. For example, the characteristic length L of plasma inhomogeneity in ICF experiments in the region of laser energy absorption usually does not exceed 100 electron mean free paths. For this reason, the interpretation of almost all experiments with a laser plasma requires the application of nonlocal transport theory.

The most significant advances in the development of nonlocal transport theory were made in the small perturbation model [9, 15, 16], for which analytic solutions to the kinetic equation can be obtained under certain assumptions and can be used for determining electron fluxes. These theories presumed that the transport processes are slow (quasi-stationary) so that transport coefficients are independent of time. In such a quasi-static approximation, the nonlocal hydrodynamic equations completely equivalent to the kinetic description of a plasma were formulated [16]. At the same time, the transient nature of transport processes may limit the application of the theory to a considerable extent [17]. In the linear theory, allowance for the effects of nonstationary transport is essentially equivalent to allowance for the ω dependence of transport coefficients, which leads to a nontrivial frequency dependence of the dielectric susceptibility. In [4, 18], these effects were taken into account for weakly collisional and collisionless plasmas. The approach developed by us here makes it possible to analyze transport properties of plasmas for any relations between the temporal, spatial, and collisional scales of plasmas.

The models of nonlocal hydrodynamics are advantageous since they simplify the description of transport processes for practical applications. Such models were developed starting from the beginning of the 1990s [19–22] and were aimed at inclusion of kinetic effects (such as Landau damping) in the comparatively simple equations of hydrodynamics. Consequently, an analytic description of transport coefficients using a consistent theory (even if it is confined to a linear approximation) would lead to further improvement of nonlocal hydrodynamic models.

In this study, the derivation of transport equations for perturbations and the expressions for dielectric susceptibility is based on solving the initial value problem for the linearized kinetic equation for plasma particles [16]. The method for solving this equation is valid for a plasma with a large ion charge $Z \gg 1$ and with arbitrary relations between the length $L = k^{-1}$ of perturbation inhomogeneity and the electron mean free path as well as between the typical temporal scale $\tau = \omega^{-1}$ of perturbation, electron collision time, and the electron mean free time (the time during which an electron traverses the distance equal to the characteristic scale of inhomogeneity, $1/kv_{Te}$, where v_{Te} is the thermal velocity of electrons). Using the expansion of the distribution function in Legendre polynomials in this approach, we sum all angular harmonics of the electron distribution function, which allows us to describe a continuous transition from the strongly collisional hydrodynamic limit to the collisionless case in the transport equations and the expression for dielectric susceptibility. The procedure for solving the initial value problem for a perturbation of the distribution function [16] is generalized to the nonstationary case. The transport equations are formulated in the form of relations between Fourier components of electron fluxes and generalized hydrodynamic forces (i.e., between the density and temperature gradients and the electric field). As a result of the transient nature of the problem, all electron transport coefficients in the (ω, k) space contain the imaginary part, which is missing in the quasi-stationary theory [9, 15, 16]. The obtained complex longitudinal dielectric susceptibility is analyzed in the entire (ω, k) region as a function of collision parameters of the plasma ($k\lambda_{ei}$ and $\omega\lambda_{ei}/v_{Te}$). The relation between the dielectric susceptibility of the plasma and nonstationary nonlocal transport coefficients is established.

2. KINETIC DESCRIPTION OF POTENTIAL PERTURBATIONS IN A COLLISIONAL PLASMA

Let us consider small potential perturbations of a homogeneous equilibrium plasma with the Maxwell distributions functions f_M^a ($a = e, i$) of electrons and ions, which are characterized by density n_a and temperature T_a. We assume that the ground state is quasi-stationary, allowing for only a slow variation of the particle temperature with time due to energy redistribution between electrons and ions as a result of collisions. In this case, the linearized kinetic equation for spatial Fourier components $\delta f_a = f_a - f_M^a$ of perturbations in the particle distribution functions has the form

$$\left(\frac{\partial}{\partial t} + i\mathbf{k}\cdot\mathbf{v}\right)\delta f_a + \frac{e_a}{m_a}\mathbf{E}\frac{\partial f_M^a}{\partial \mathbf{v}} = C_{ab}[\delta f_a, f_b] + C_{ab}[f_a, \delta f_b] + C_{aa}[\delta f_a, f_a] + C_{aa}[f_a, \delta f_a], \quad (1)$$

where C_{ab} and C_{aa} are the Landau collision integrals for particles of the same species with a charge e_a and a mass m_a.

Carrying out the unilateral Fourier transformation in time and expanding the distribution function δf_a in Legendre polynomials $P_l(\theta)$,

$$\delta f_a = \sum_{l=0}^{\infty} f_l^a P_l(\theta), \quad C_{ab} = \sum_{l=0}^{\infty} C_{ab}^l P_l(\theta), \tag{2}$$

we obtain an infinitely large system of equations for the (ω, k) Fourier components of angular harmonics of distribution functions f_l^a:

$$\begin{gathered} -i\omega f_l^a + ikv\frac{l}{2l-1}f_{l-1}^a + ikv\frac{l+1}{2l+3}f_{l+1}^a \\ -C_{aa}^l - C_{ab}^l = S_l^a. \end{gathered} \tag{3}$$

Here, the collision integrals C_{ab}^l (both for $b = a$ and for $b \neq a$) written using the Rosenbluth potentials [8] have the form

$$\begin{aligned} \frac{C_{ab}^l}{\nu_{ab}(v)} &= \frac{l(l+1)}{6} f_l^a (I_2^0 - 3I_0^0 - 2J_{-1}^0) \\ &+ \frac{v}{3}\frac{\partial}{\partial v}\left(v\frac{\partial f_l^a}{\partial v}(I_2^0 + J_{-1}^0)\right) \\ &+ \frac{m_a}{m_b} v \frac{\partial}{\partial v}(f_l^a I_0^0) + \frac{m_a}{m_b}\frac{4\pi v^3}{n_b} f_M^a f_l^b + v\frac{\partial f_M^a}{\partial v} \\ &\times \frac{l\delta J_{-1-l}^l - (l+1)\delta I_l^l}{2l+1}\left(1 - \frac{m_a}{m_b}\right) \\ &+ \frac{v^2}{2(2l+1)}\frac{\partial^2 f_M^a}{\partial v^2}\left(\frac{l(1-l)}{2l-1}(\delta I_l^l + \delta J_{1-l}^l)\right. \\ &\left. + \frac{(1+l)(2+l)}{2l+3}(\delta I_{l+2}^l + \delta J_{-1-l}^l)\right) \\ &+ \frac{v}{2(2l+1)}\frac{\partial f_M^a}{\partial v}\left(\frac{(l^2+3l-2)\delta I_l^l + l(l-1)\delta J_{1-l}^l}{2l-1}\right. \\ &\left. - \frac{(l+1)(l+2)\delta I_{l+2}^l + (l^2-l-4)\delta J_{-1-l}^l}{2l+3}\right), \end{aligned} \tag{4}$$

where $\nu_{ab}(v) = 4\pi n_b(e_a e_b)^2\Lambda_{ab}/m_a^2 v^3$ is the velocity-dependent frequency of collisions of particles of species a with particles of species b, Λ_{ab} is the Coulomb logarithm, and

$$\begin{aligned} \{I_m^0; \delta I_m^n\} &= \frac{4\pi}{n_b v^m}\int_0^v \{f_M^b; f_n^b\} v^{m+2} dv, \\ \{J_m^0; \delta J_m^n\} &= \frac{4\pi}{n_b v^m}\int_v^\infty \{f_M^b; f_n^b\} v^{m+2} dv \end{aligned} \tag{5}$$

are unperturbed Rosenbluth potentials J_m^0 and their perturbations δJ_m^n, which can be defined from standard expressions [8].

Under the assumption that the initial perturbation $\delta f_a(t = 0)$ of the distribution function is determined by the perturbed Maxwell distribution (i.e., characterized by the initial perturbations of density $\delta n_a(0)$ and temperature $\delta T_a(0)$ [16]),

$$\begin{aligned} &\delta f(v, t = 0) \\ &= \left[\frac{\delta n_a(0)}{n_a} + \frac{\delta T_a(0)}{T_a}\left(\frac{v^2}{2v_{Ta}^2} - \frac{3}{2}\right)\right] f_M^a(v), \end{aligned} \tag{6}$$

the sources in kinetic equation (3) (right-hand sides) appear only for the first two angular harmonics of the distribution function ($S_l^a = 0$ for $l \geq 2$) and are determined by the initial perturbations of the distribution function and by the (ω, k) Fourier component of the electric field, namely, $S_0 = \delta f(t = 0)$ and $S_1^a = (e_a E/T_a) v f_M^a$.

Taking two moments in velocity of Eq. (3) with $l = 0$ for perturbed density δn_a and temperature δT_a,

$$\begin{aligned} \delta n_a &= 4\pi\int_0^\infty dv\, v^2 f_0^a, \\ \delta T_a &= \frac{4\pi m_a}{3n_a}\int_0^\infty dv\, v^2(v^2 - 3v_{Ta}^2) f_0^a \end{aligned} \tag{7}$$

($v_{Ta} = \sqrt{T_a/m_a}$ is the thermal velocity), we obtain the number-of-particles and energy conservation laws,

$$\frac{\partial \delta n_a}{\partial t} + ikn_a u_a = 0, \tag{8}$$

$$\begin{aligned} &\frac{\partial \delta T_a}{\partial t} + ik\frac{2}{3n_a}(q_a + n_a u_a T_a) \\ &= \frac{4\pi m_a}{3n_a}\int dv\, v^4 C_{ab}^0 \equiv \left.\frac{\partial \delta T_a}{\partial t}\right|_{ab}, \end{aligned} \tag{9}$$

where u_a is the average velocity of particles and q_a is their heat flux:

$$\begin{aligned} u_a &= \frac{4\pi}{3n_a}\int dv\, v^3 f_1^a, \\ q_a &= \frac{2\pi T_a}{3}\int dv\, v^3\left(\frac{v^2}{v_{Ta}^2} - 5\right) f_1^a. \end{aligned} \tag{10}$$

Here and everywhere below, the thermal energy of particles is given in the normalization corresponding to the Boltzmann constant equal to unity. The right-hand side in Eq. (9) describes the collisional energy exchange between the species of particles:

$$\begin{aligned} \int dv\, v^4 C_{ab}^0 &= 2\int dv\, \nu_{ab} v^4 \\ &\times\left\{ f_M^a\left(\delta J_{-1}^0 - \frac{m_a}{m_b}\delta I_0^0\right) + f_a^0\left(J_{-1}^0 - \frac{m_a}{m_b} I_0^0\right)\right\}. \end{aligned} \tag{11}$$

Accordingly, for the moment of u_a, we obtain from Eq. (3) with $l = 1$ the law of momentum conservation

$$\begin{aligned} n_a m_a \frac{\partial u_a}{\partial t} + ik(\delta n_a T_a + \delta T_a n_a) + ik\Pi_\parallel^a - e_a n_a E \\ = \frac{4\pi m_a}{3}\int dv\, v^3 C_{ab}^1 \equiv n_a m_a \left.\frac{\partial u_a}{\partial t}\right|_{ab}, \end{aligned} \tag{12}$$

where the contributions from collision (friction) are described as follows:

$$\begin{aligned} \int dv\, v^3 C_{ab}^1 &= \left(1 + \frac{m_a}{m_b}\right) \\ &\times \int dv\, v^3 \nu_{ab}(f_M^a \delta J_{-2}^1 - f_1^a I_0^0). \end{aligned} \tag{13}$$

Quantity $\Pi_\parallel^a$ in expression (12) is given by expression

$$\Pi_\parallel^a = \frac{8\pi m_a}{15}\int dv\, v^4 f_2^a \tag{14}$$

as the longitudinal component of the stress tensor.

Further simplification is associated with expansion of the electron–ion and ion–electron collision integrals in the characteristic velocity of ions and the electron velocity in corresponding equations (3). This gives the following expressions for these collision integrals:

$$\begin{aligned} \frac{C_{ei}^l}{\nu_{ei}(v)} &= \frac{1}{2} l(l+1) f_l^e + \frac{4\pi v}{n_i} \\ &\times \frac{\partial}{\partial v}\left\{\frac{m_e}{m_i}\left(f_l \int_0^v dv' v'^2 f_M^{i\prime} + \delta_{l0} f_M^e \int_0^v dv' v'^2 f_0^{i\prime}\right)\right. \\ &\left. + \frac{1}{3v}\left(\frac{\partial f_l^e}{\partial v}\int_0^v dv' v'^4 f_M^{i\prime} + \delta_{l0}\frac{\partial f_M^e}{\partial v}\int_0^v dv' v'^4 f_0^{i\prime}\right)\right\} \\ &- \delta_{l1} u_i \frac{\partial f_M^e}{\partial v}, \end{aligned} \tag{15}$$

$$\begin{aligned} \frac{C_{ie}^l}{\nu_{ie}(v)} &= -\frac{v}{3n_e v_{Te}}\sqrt{\frac{2}{\pi}} l(l+1) f_l^i \\ &+ \frac{4\pi v^3}{3 v_{Ti}^2}\frac{m_i}{m_e}\delta_{l1}\int_0^\infty dv' f_1^{e\prime} \\ &+ \frac{v}{3n_e}\frac{\partial}{\partial v}\left\{\sqrt{\frac{\pi}{2}}\left(\frac{m_i v^3}{m_e v_{Te}^3} f_l^i + \frac{v^2}{v_{Te}}\frac{\partial f_l^1}{\partial v}\right)\right. \\ &\left. + 4\pi\delta_{l0} f_M^i v^3\left(f_0^e(0) - \frac{1}{v_{Ti}^2}\int_0^\infty dv' v' f_0^{e\prime}\right)\right\}, \end{aligned} \tag{16}$$

in these expressions, primed distribution functions correspond to argument v'. The above collision integrals give the conservation laws for the number of particles, momentum, and energy. The energy exchange between electrons and ions (cooling) in Eq. (9) for electron temperature perturbations associated with collisions is described by the expression

$$\begin{aligned} \left.\frac{\partial \delta T_e}{\partial t}\right|_{ei} &= 2\nu_{ei}^T \frac{m_e}{m_i}\left\{\frac{\delta n_i}{n_i}(T_i - T_e) + \delta T_i\right. \\ &\left. + \frac{4\pi}{n_e}\sqrt{\frac{\pi}{2}}\left(v_{Te}^3 f_0^e(0) T_i - T_e v_{Te}\int dv\, v f_0^e\right)\right\}, \end{aligned} \tag{17}$$

where $\nu_{ei}^T = 2\nu_{ei}(v_{Te})/(3\sqrt{2\pi})$. The energy balance equation for ions contains the same contribution with the opposite sign (heating). Accordingly, the e–i collisions give a contribution to the momentum conservation law (12),

$$\begin{aligned} \left.\frac{\partial u_e}{\partial t}\right|_{ei} &= \nu_{ei}^T u_i - \frac{1}{m_e n_e}\mathcal{R}_{ie}, \\ \mathcal{R}_{ie} &= \frac{4\pi m_e}{3}\int dv\, v^3 \nu_{ei}(v) f_1^e, \end{aligned} \tag{18}$$

which takes into account the friction force $\mathcal{R}_{ie}$ and the ion recoil $\nu_{ei}^T u_i$.

Since we are interested only in kinetic effects for electrons, we seek a solution to Eq. (3) only for the electron distribution function. Assuming that ions have a large degree of ionization $Z \gg 1$, we can disregard the electron–electron (*e*–*e*) collision in the equations for higher harmonics of the electron distribution function, retaining only C_{ee}^0 in the equation for its symmetric part. We also ignore the contributions on the order of $\sim m_e/m_i$ in C_{ei}, which are responsible for the slow energy transfer from electrons to ions, which is significant only over large temporal scales. This is well justified, for example, for a laser plasma. Thus, for the electron–ion collision integral, we will use the expression

$$C_{ei}^l = -\frac{l(l+1)}{2}\nu_{ei} f_l^e + \delta_{l1}\nu_{ei}\frac{v}{v_{Te}^2} f_M^e u_i, \qquad (19)$$

assuming that the mean ion velocity u_i appearing in C_{ei} is a preset quantity. An analogous system of equations for f_l^e was considered in [16] using a reference frame in which ions are at rest. However, in contrast to our approach, the authors of [16] used a quasi-stationary approximation imposing the limitation of a low rate of transport processes. Here, this constraint does not exist.

The equations for higher ($l > 1$) angular harmonics of distribution function (3) are eliminated by introducing modified collision frequency ν_1 [16, 23], for which we can write the recurrence relation [23],

$$\begin{aligned}\nu_l(v) &= -i\omega + \frac{1}{2}l(l+1)\nu_{ei}(v)\\ &+ \frac{(l+1)^2}{4(l+1)^2 - 1}\frac{k^2v^2}{\nu_{l+1}(v)},\end{aligned} \qquad (20)$$

this makes it possible to write the following equation for the first angular harmonic:

$$f_1^e = -\frac{ikv}{\nu_1(v)} f_0^e + \frac{1}{\nu_1(v)}\frac{\partial f_M^e}{\partial v}\left(\frac{eE}{m_e} - \nu_{ei}u_i\right). \qquad (21)$$

Here, the symmetric part of perturbations of the electron distribution function f_0^e satisfies the kinetic equation

$$\begin{aligned}&\left(-i\omega + \frac{k^2v^2}{3\nu_1}\right) f_0^e - C_{ee}[f_0^e]\\ &= \frac{ikv^2}{3\nu_1}\frac{eE}{T_e} f_M^e - iku\frac{v^2\nu_{ei}}{3v_{Te}^2\nu_1} f_M^e + \delta f^e(v, t = 0)\end{aligned} \qquad (22)$$

with the initial perturbation $f^e(v, t = 0)$ defined by relation (6).

Following [16], we present the general solution to Eq. (22) using the basis functions ψ^A,

$$\begin{aligned}f_0^e &= i\frac{eE}{kT_e} f_M^e + \left(\frac{\delta n_e(0)}{n_e} - \omega\frac{eE}{kT_e}\right)\psi^N f_M^e\\ &+ \frac{3}{2}\frac{\delta T_e(0)}{T_e}\psi^T f_M^e - iku_i\psi^R f_M^e,\end{aligned} \qquad (23)$$

which satisfy the kinetic equation with various sources S_A ($A = N, T, R$),

$$\left(-i\omega + \frac{k^2v^2}{3\nu_1}\right)\psi^A = (f_M^e)^{-1} C_{ee}[f_M^e\psi^A] + S_A, \qquad (24)$$

where $S_N = 1$, $S_T = v^2/3v_{Te}^2 - 1$, and $S_R = v^2\nu_{ei}/3v_{Te}^2\nu_1$. We also introduce the moments J_B^A of basis function [16],

$$J_B^A = \frac{4\pi}{n_e}\int_0^\infty v^2 dv\,\psi^A f_M^e S_B, \qquad (25)$$

which are symmetric to transposition of the upper and lower indices.

We can eliminate the initial perturbations from Eq. (23) by calculating the first two moments of δn_e and δT_e (7), which leads to the following relations between the initial perturbations ($\delta n_e(0)$ and $\delta T_e(0)$) and current perturbations (δn_e and δT_e):

$$\begin{aligned}\frac{\delta n_e}{n_e} &= i\frac{eE}{kT_e} + \left(\frac{\delta n_e(0)}{n_e} - \omega\frac{eE}{kT_e}\right) J_N^N\\ &+ \frac{3}{2}\frac{\delta T_e(0)}{T_e} J_N^T - iku_i J_N^R,\\ \frac{\delta T_e}{T_e} &= \left(\frac{\delta n_e(0)}{n_e} - \omega\frac{eE}{kT_e}\right) J_T^N\\ &+ \frac{3}{2}\frac{\delta T_e(0)}{T_e} J_T^T - iku_i J_T^R.\end{aligned} \qquad (26)$$

Accordingly, having eliminated the initial perturbations $\delta n_e(0)$ and $\delta T_e(0)$ from Eq. (23), we can derive the expression for the symmetric component of the electron distribution function f_0^e proceeding from the lower hydrodynamic moments

$$\begin{aligned}f_0^e &= i\frac{eE}{kT_e} f_M^e + \left(\frac{\delta n_e}{n_e} - i\frac{eE}{kT_e}\right)\frac{J_T^T\psi^N - J_T^N\psi^T}{D_{NT}^{NT}} f_M^e\\ &+ \frac{\delta T_e}{T_e}\frac{J_N^N\psi^T - J_N^T\psi^N}{D_{NT}^{NT}} f_M^e\\ &- iku_i\left(\psi^R - \frac{D_{NT}^{RT}}{D_{NT}^{NT}}\psi^N - \frac{D_{NT}^{NR}}{D_{NT}^{NT}}\psi^T\right) f_M^e,\end{aligned} \qquad (27)$$

where $D_{AB}^{CD} = J_A^C J_B^D - J_A^D J_B^C$. The resulting expression together with Eq. (21) for f_1^e makes it possible to formulate the linear theory of electron response of plasma and transport.

3. NONLOCAL HYDRODYNAMICS FOR ELECTRON PERTURBATIONS

In our previous discussion, we realized the approach that has made it possible to express the electron distribution function in terms of its lower moments. This solves the problem of closing the chain of equations for hydrodynamic moments. The simplest closure of this type is well known in the strongly collisional case, where the Chapman–Enskog or Grad methods are used. However, this imposes severe constraints on the relation between the electron mean free path λ_{ei} and the characteristic scale L of perturbation inhomogeneity in a plasma [24], $\lambda_{ei}/L < 0.06/\sqrt{Z}$. Consequently, the classical theory cannot be used for describing experiments on the interaction of laser radiation with matter in fusion studies, in which small-scale perturbations are of practical interest. Considerable expansion of the range of application of hydrodynamic equations, which are convenient for describing plasmas, was achieved in the framework of nonlocal hydrodynamics [16] formulated for slow processes in the quasi-static approximation. Here, we solve the problem of generalization of nonlocal hydrodynamics to the case of rapidly varying processes in a plasma.

A. Nonlocal Transport Equations

The first three moments of kinetic equation (1) lead to the equations of continuity, motion, and energy balance for electrons,

$$\begin{aligned} &\frac{\partial \delta n_e}{\partial t} + n_e i\mathbf{k}\cdot\mathbf{u}_e = 0, \\ &\frac{\partial \mathbf{u}_e}{\partial t} = -\frac{e}{m_e}\mathbf{E}^* + \frac{1}{m_e n_e} i\mathbf{k}\cdot\hat{\Pi}^e - \frac{1}{m_e n_e}\mathbf{R}_{ie}, \\ &\frac{\partial \delta T_e}{\partial t} + \frac{2}{3n_e} i\mathbf{k}\cdot\mathbf{q}_e + \frac{2}{3}T_e i\mathbf{k}\cdot\mathbf{u}_e = 0, \end{aligned} \tag{28}$$

where $\mathbf{u}_e$ is the electron drift velocity and $\mathbf{E}^*$ is the Fourier component of the effective electric field,

$$\mathbf{E}^* = \mathbf{E} + i\mathbf{k}\frac{T_e}{e}\left(\frac{\delta n_e}{n_e} + \frac{\delta T_e}{T_e}\right). \tag{29}$$

Following previous publications [16], we have introduced the generalized friction force $\mathbf{R}_{ie} = \mathcal{R}_{ie} - m_e n_e \nu_{ei}^T \mathbf{u}_i$. It should be noted that the equation of motion for electrons (second equation) can be used for defining the stress tensor

$$i\mathbf{k}\hat{\Pi}^e = \mathbf{R}_{ei} + en_e\mathbf{E}^* - i\omega m_e n_e \mathbf{u}_e,$$

while the remaining two equations of system (28) are completely equivalent to system (26).

Integrating Eqs. (10) and (21), we can represent electric current $\mathbf{j} = en_e(\mathbf{u}_i - \mathbf{u}_e)$ and the heat flux in the standard form

$$\begin{aligned} \mathbf{j} &= \sigma\mathbf{E}^* + \alpha i\mathbf{k}\delta T_e + \beta_j e n_e \mathbf{u}_i, \\ \mathbf{q}_e &= -\alpha T_e \mathbf{E}^* - \chi i\mathbf{k}\delta T_e - \beta_q n_e T_e \mathbf{u}_i, \end{aligned} \tag{30}$$

while for the friction force we obtain

$$\mathbf{R}_{ie} = (\beta_j - 1)n_e e\mathbf{E}^* + \beta_q n_e i\mathbf{k}\delta T_e - \nu_{ei}^T\beta_r m_e n_e \mathbf{u}_i. \tag{31}$$

Here, we have introduced nonlocal nonstationary transport coefficients in the (ω, k) space: electrical conductivity σ, thermoelectric coefficient α, thermal diffusivity χ, and ion convection transport coefficient $\beta_{j,q,r}$:

$$\begin{aligned} \sigma &= \frac{e^2 n_e}{k^2 T_e}\left(\frac{J_T^T}{D_{NT}^{NT}} + i\omega\right), \\ \alpha &= -\frac{e n_e}{k^2 T_e}\left(\frac{J_T^N + J_T^T}{D_{NT}^{NT}} + i\omega\right), \\ \beta_j &= 1 - \frac{D_{NT}^{RT}}{D_{NT}^{NT}}, \\ \chi &= \frac{n_e}{k^2}\left(\frac{2J_T^N + J_T^T + J_N^N}{D_{NT}^{NT}} + i\frac{5}{2}\omega\right), \\ \beta_q &= \frac{D_{NT}^{RT} + D_{NT}^{RN}}{D_{NT}^{NT}}, \\ \beta_r &= 1 + k^2 v_{Te}\lambda_{ei} \\ &\times [J_R^R - (1-\beta_j)(J_R^N + J_R^T) + \beta_q J_R^T] \\ &- (2\pi)^{3/2}\frac{v_{Te}}{n_e}\int_0^\infty \frac{dv\, v\, \nu_{ei}(v)}{\nu_1(v)} f_M^e. \end{aligned} \tag{32}$$

All electric transport coefficients are complex functions which, being presented in dimensionless form, can be parametrized using $k\lambda_{ei}$, ω/ν_{ei}^T and Z using classical expressions. In the static limit $\omega = 0$, nonlocal transport coefficients (32) were analyzed in [16]. The range of applicability of the static transport coefficients in the classical strongly collisional limit ($k\lambda_{ei} < 0.06/\sqrt{Z}$) in a plasma with a high degree of ionization of ions, $Z \gg 1$, is determined by the smallness of frequency as compared to the frequency of the *e*–*i* collisions, $\omega \ll \nu_{ei}^T$ [8]. The coefficients themselves are in fact determined by quantity ν_{ei}^T, and the effects associated with electron–electron collision are small corrections of order $O(Z^{-1})$ [13, 14]. The applicability limits of this classical limits are determined by the electron energy delocalization length $\lambda_\epsilon = \sqrt{Z}\,\lambda_{ei}$ [11, 12], which defines the spatial scale for which the relaxation rate associated with *e*–*e* collisions becomes equal to the rate of spatial transport in kinetic equation (22). In the vicinity of this boundary, *e*–*e* collisions begin to strongly affect the transport coefficients, modifying the symmetric part of the distribution function, which in turn determines the anisotropic correction (see Eq. (21)) as well as electron fluxes. This narrows the applicability limits for the static approximation for transport coefficients, which is now determined by the relation between ω and the *e*–*e* collision frequency. At the same time, with increasing $k\lambda_{ei}$, as a result of the electron redistribution associated with spatial transport, subthermal electrons (i.e., electrons with low velocities $v \lesssim v_{Te}$) begin to determine perturbation of the symmetric part of the distribution function, thus effectively increasing the frequency of *e*–*e* collisions [9]. For example, for $k\lambda_{ei} \gg \sqrt{Z}$, their characteristic velocities $v^* \sim v_{Te}/(Zk^2\lambda_{ei}^2)^{1/7}$ become noticeably lower than thermal velocities [9, 16]. Thus, the range of application of static approximation for transport coefficients for moderate gradients $0.06/\sqrt{Z} < k\lambda_{ei} < 6Z^{2/3}$ are determined by the condition $\omega \ll \nu_{ee}^T,\ \nu_{ei}^T(k\lambda_{ei})^{4/7}/Z^{5/7}$ [9, 16], where $\nu_{ee}^T = 2\nu_{ee}(v_{Te})/(3\sqrt{2\pi})$. For higher gradients in the region of $k\lambda_{ei} \sim 6Z^{2/3}$, all angular harmonics must be taken into account for obtaining a correct description of the transition to the collisionless limit. This explains the sharp transition to the region in which the applicability of the static approximation to transport coefficient is violated. In the collisionless range of wavenumbers, $k\lambda_{ei} > 6Z^{2/3}$, this region is defined in the standard manner as $\omega \ll kv_{Te}$. Henceforth, we will define the applicability region of the static approximation for the transport coefficients

$$\omega \ll \begin{cases} \nu_{ei}^T, & k\lambda_{ei} > 0.06/\sqrt{Z}, \\ \nu_{ee}^T,\ \nu_{ei}^T(k\lambda_{ei})^{4/7}/Z^{5/7}, & 0.06/\sqrt{Z} < k\lambda_{ei} < 6Z^{2/3}, \\ kv_{Te}, & k\lambda_{ei} > 6Z^{2/3}, \end{cases} \tag{33}$$

as the quasi-static limit of our nonlocal theory. It should be noted that the above-mentioned sharp transition to the collisionless kinetic limit is expressed in that expression (33) formally experiences a jump for $k\lambda_{ei} \sim 6Z^{2/3}$.

B. Nonlocal Transport Coefficients

The region of application of the classical (local) theory is well defined in the quasi-static limit (33) [24]. This region expands with increasing ω since a simple exact solution to the kinetic equation [8] exists for $|\omega + i\nu_{Te}^T| \gg kv_{Te}$ if we disregard *e*–*e* collisions. Our exact solution shows that nonlocal effects are insignificant for $k\lambda_{ei} < 0.06/\sqrt{Z}$, $0.1\omega/\nu_{ei}^T$. In this case, Eq. (24) can be solved by expanding the basis functions using the first two Laguerre polynomials $\psi^A = C_0^A + C_1^A(v^2/3v_{Te}^2 - 1)$. Then the effective frequency of collisions satisfies the following approximate expression: $\nu_1 = \nu_{ei} - i\omega$. As a result, we obtain the transport coefficients in quadratures [8]:

$$\begin{aligned} \frac{\sigma}{\sigma_{SH}} &= \frac{1}{48}\int_0^\infty dx x^6 Q(x), \\ \frac{\alpha}{\alpha_{SH}} &= \frac{1}{144}\int_0^\infty dx x^6 (x^2-5) Q(x), \\ \frac{\chi}{\chi_{SH}} &= \frac{1}{1200}\int_0^\infty dx x^6 (x^2-5)^2 Q(x), \\ \beta_r &= 1 - \int_0^\infty dx Q(x), \\ \beta_j &= 1 - \sqrt{\frac{2}{9\pi}}\int_0^\infty dx x^3 Q(x), \\ \beta_q &= \sqrt{\frac{1}{18\pi}}\int_0^\infty dx x^3 (5-x^2) Q(x), \end{aligned} \tag{34}$$

where we have introduced the notation

$$Q(x) = \frac{\nu_{ei}^T x \exp(-x^2/2)}{\nu_{ei}^T - i\sqrt{2/9\pi}\,\omega x^3}, \quad x = v/v_{Te},$$

and

$$\sigma_{SH} = \frac{32 n_e e^2}{3\pi m_e \nu_{ei}^T}, \quad \alpha_{SH} = \frac{16 n_e e}{\pi m_e \nu_{ei}^T},$$

$$\chi_{SH} = \frac{200 n_e v_{Te} \lambda_{ei}}{3\pi}$$

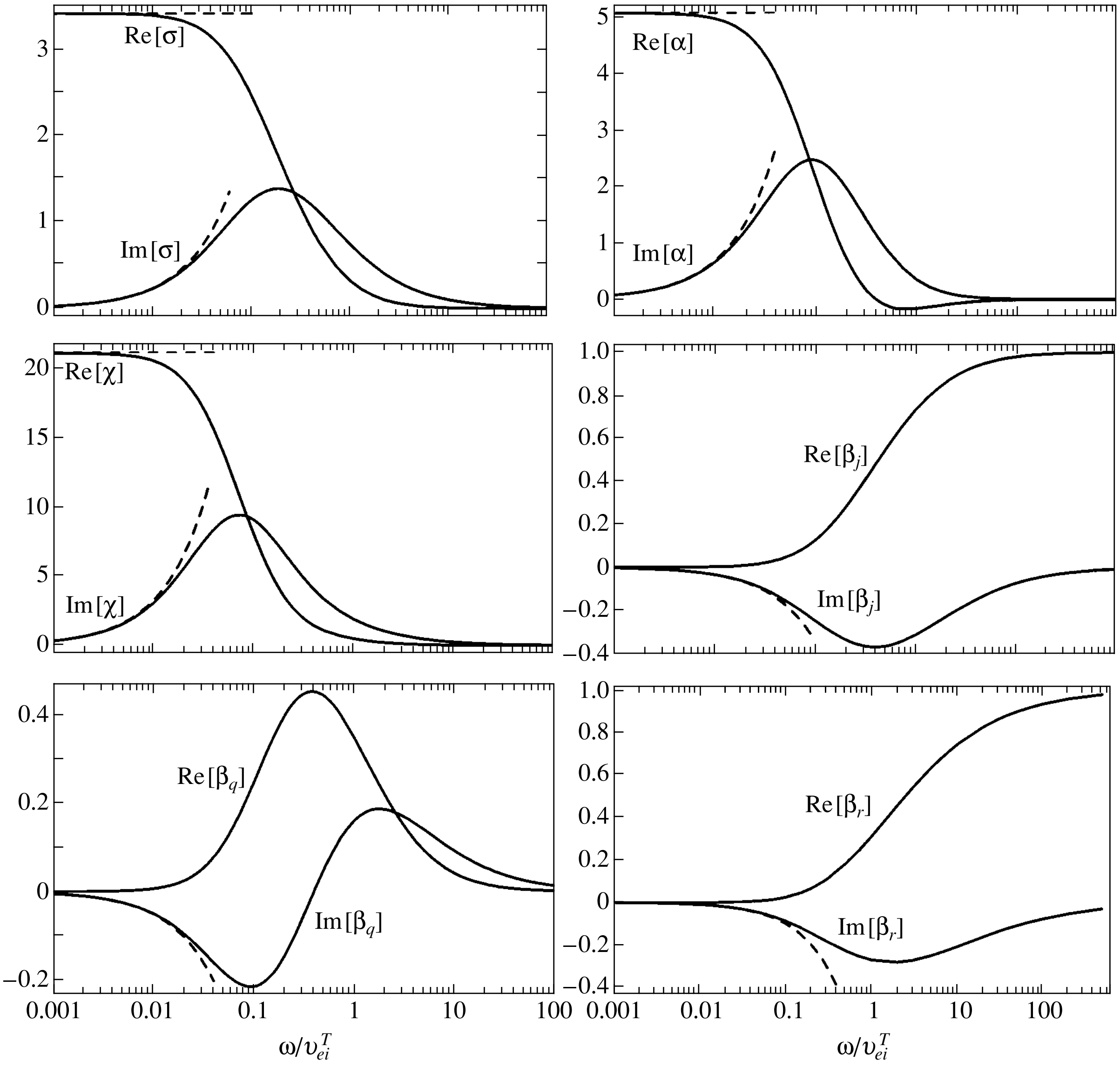


Fig. 1. Dependence of the real and imaginary parts of transport coefficients σ, α, χ, and $\beta_{j,\,q,\,r}$ on ω/ν_{ei}^T in the long-wave limit $k\lambda_{ei} < 0.06/\sqrt{Z}$, $0.1\omega/\nu_{ei}^T$. Dashed curves correspond to the static limit (35).

correspond to the classical transport coefficients [13, 14]. These expressions for transport coefficients are independent of the wavenumber and correspond to the local limit, including the hydrodynamic limit. Figure 1 illustrates their dependence on frequency.

In the limit of strong collisions and low frequencies $\omega \ll \nu_{ei}^T$, expressions (34) lead to classical real-valued transport coefficients [13] with small imaginary corrections

$$\frac{\sigma}{\sigma_{SH}} = 1 + i\frac{105}{16}\frac{\omega}{\nu_{ei}^T},$$

$$\frac{\alpha}{\alpha_{SH}} = 1 + i\frac{105}{8}\frac{\omega}{\nu_{ei}^T}, \quad \frac{\chi}{\chi_{SH}} = 1 + i\frac{609}{40}\frac{\omega}{\nu_{ei}^T},$$
$$\beta_j = -i\frac{32}{3\pi}\frac{\omega}{\nu_{ei}^T}, \quad \beta_q = -i\frac{32}{2\pi}\frac{\omega}{\nu_{ei}^T}, \tag{35}$$
$$\beta_r = -i\frac{\omega}{\nu_{ei}^T}.$$

With increasing ω, the real part of the classical transport coefficients σ, α, and χ decrease monotonically, while the imaginary part increases, attaining its maximal value for $\omega \sim 0.1\nu_{ei}^T$ and then decreases (see Fig. 1). In the limit of strong collisions, the ion convective transport coefficients are negligibly small; these coefficients appear only due to nonstationary nature of the process since it can be seen directly that they are proportional to ω. In the high-frequency limit $\omega \gg \nu_{ei}^T$, coefficients β_j and β_r with small imaginary parts tend to unity, while coefficient β_q is small in absolute value, its real part being smaller than the imaginary type. In the same limit, transport coefficients σ, α, and χ become purely imaginary with small real corrections:

$$\sigma = \frac{ie^2 n_e}{m_e\omega}\left(1 - i\frac{\nu_{ei}^T}{\omega}\right),$$

$$\chi = \frac{iT_e n_e}{m_e\omega}\left(\frac{5}{2} - i\frac{13\nu_{ei}^T}{4\omega}\right),$$

$$\alpha = \frac{ien_e}{m_e\omega}\left(\frac{5}{2}\left(\frac{\pi^8(\nu_{ei}^T)^5}{36\omega^5}\right)^{1/6} + i\frac{3\nu_{ei}^T}{2\omega}\right), \tag{36}$$

$$\beta_j = 1 - i\frac{\nu_{ei}^T}{\omega}, \quad \beta_q = -i\frac{3\nu_{ei}^T}{2\omega},$$

$$\beta_r = 1 - \left(\frac{9\pi(\nu_{ei}^T)^2}{2\omega^2}\right)^{1/3}\frac{\pi}{3\sqrt{3}}(1 + i\sqrt{3}).$$

With increasing collision parameter $k\lambda_{ei}$, the nature of the frequency dependence of transport coefficient also changes. For example, coefficients α and χ characterize the essentially nonmonotonic frequency dependence, which can be clearly seen in Fig. 2. For example, the real part of the thermal diffusivity first increases with frequency ω as compared to the static case, and then decreases for $\omega/\nu_{ei}^T > 1$. Accordingly, the imaginary part of the thermal diffusivity is first negative and decreases to its minimal value; then it begins to increase, changes sign, attains its maximal value, and decreases again. Both the imaginary and real part of the thermal diffusivity have a maximum for $\omega/\nu_{ei}^T \sim 1$ at $k\lambda_{ei} = 1$ (see Fig. 2). A still more complex frequency dependence appears for the thermoelectric coefficient α whose imaginary and real parts have three points of inflection each. The real part of α reverses its sign upon an increase in ω. The above arguments indicate a nontrivial manifestation of the Peltier effect in a hot plasma.

Relatively simple equations for transport coefficients can be obtained in the (ω, k) region, in which the *e–e* collisions make an insignificant contribution [6], i.e., for

$$\omega \gg \nu_{ee}^T, \quad \nu_{ei}^T(k\lambda_{ei})^{4/7}/Z^{5/7}.$$

In this case, obtaining solutions for the basis distribution functions in the form

$$\psi^A = 3\nu_1 S_A/(k^2 v^2 - 3i\omega\nu_1),$$

we arrive at the following expressions for their moments:

$$J_N^N = \frac{3}{kv_{Te}}\int_0^\infty dx W(x),$$

$$J_N^T = \frac{1}{kv_{Te}}\int_0^\infty dx(x^2 - 3)W(x),$$

$$J_T^T = \frac{1}{3kv_{Te}}\int_0^\infty dx(x^2 - 3)^2 W(x),$$

$$J_R^N = \frac{3\nu_{ei}^T}{kv_{Te}}\sqrt{\frac{\pi}{2}}\int_0^\infty \frac{dxW(x)}{x\nu_1(x)}, \tag{37}$$

$$J_R^T = \frac{\nu_{ei}^T}{kv_{Te}}\sqrt{\frac{\pi}{2}}\int_0^\infty \frac{dxW(x)(x^2 - 3)}{x\nu_1(x)},$$

$$J_R^R = \frac{3\pi(\nu_{ei}^T)^2}{2kv_{Te}}\int_0^\infty \frac{dxW(x)}{x^2\nu_1^2(x)},$$

where

$$W(x) = \frac{\sqrt{2/\pi}kv_{Te}\exp(-x^2/2)}{k^2v_{Te}^2/\nu_1(x) - 3i\omega/x^2}.$$

All transport coefficients (32) can easily be calculated in terms of these moments.

In the collisionless kinetic limit $k\lambda_{ei} \gg 1$, we have β_j, $\beta_r = 1$, $\beta_q = 0$, while the remaining coefficients are functions of parameter $p = \omega/kv_{Te}$ and can be obtained using $\nu_1 = kv_{Te}h_1$ as the effective frequency in Eqs. (37), where $h_{l-1} = -ip + x^2l^2/(4l^2 - 1)h_l$ (cf. Eq. (20)). In this case, we can propose a simple approximate formula,

$$h_1(x, p) = i(\pi(p - \sqrt{p^2 - x^2})/6 - p), \tag{38}$$

for h_1, which coincides with the exact solution to within 1%. Substituting expressions (37) calculated in this

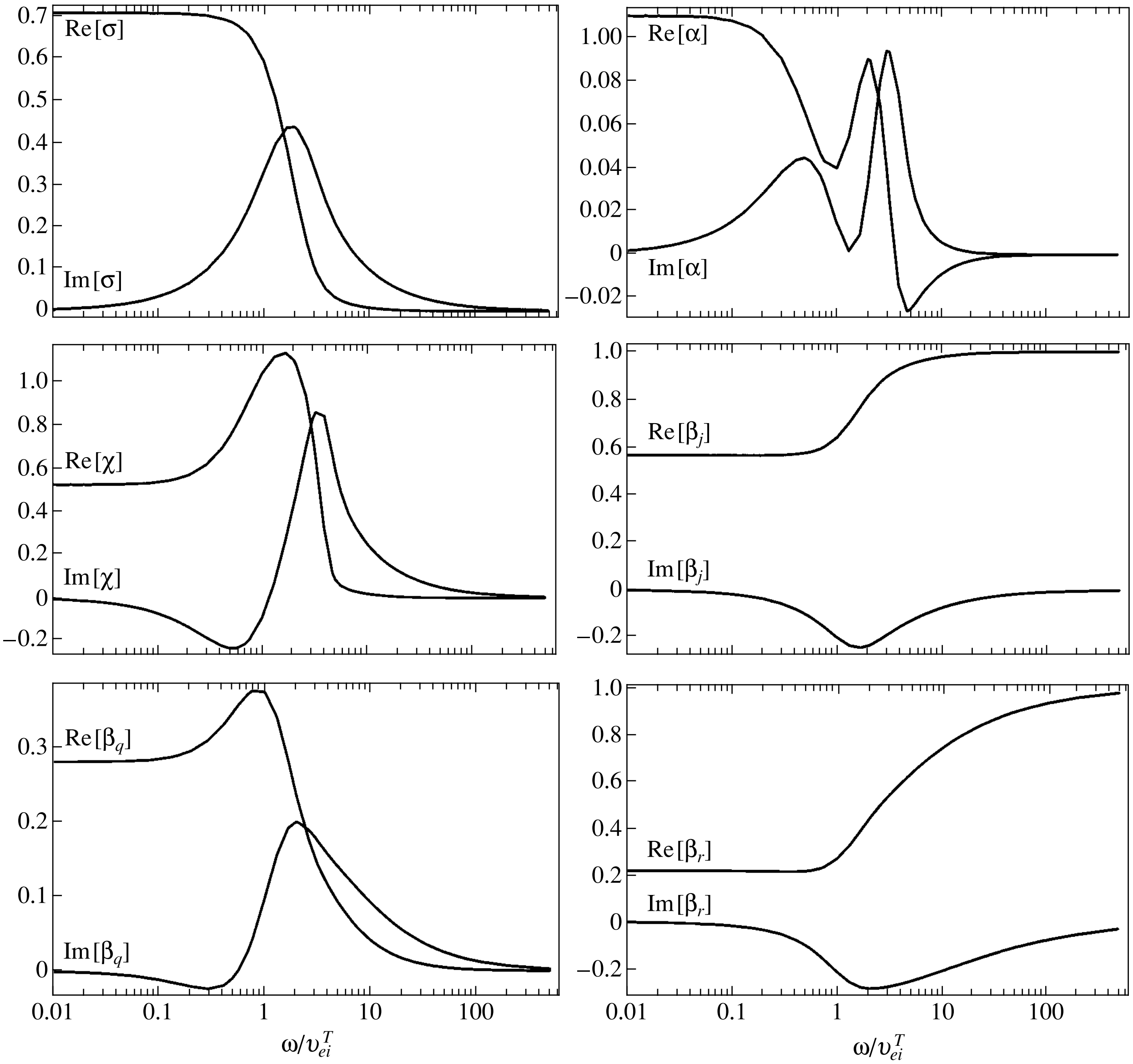


Fig. 2. Dependence of the real and imaginary parts of transport coefficients σ, α, χ, and $\beta_{j,\,q,\,r}$ on ω/ν_{ei}^T for $k\lambda_{ei} = 1$.

way into Eqs. (32), we obtain

$$\sigma = \frac{e^2 n_e v_{Te}}{kT_e} \times\left(\frac{1}{\Delta}\int_0^\infty\left(\frac{x^4}{3} - 2x^3 + 3\right)W(x)dx + ip\right),$$

$$\alpha = \frac{e n_e v_{Te}}{kT_e}\left(\frac{1}{\Delta}\int_0^\infty\left(x^2 - \frac{x^4}{3}\right)W(x)dx - ip\right), \tag{39}$$

$$\chi = \frac{n_e v_{Te}}{k}\left(\frac{1}{\Delta}\int_0^\infty\frac{x^4}{3}W(x)dx + i\frac{5}{2}p\right),$$

where the following notation is used:

$$\Delta = \int x^4 W(x)dx\int W(x)dx - \left(\int x^2 W(x)dx\right)^2.$$

It should be noted that expressions for collisionless transport coefficients were obtained in [18] using a representation differing from (39) using explicit summation of an infinite series. At the same time, the collisionless transport coefficients can be calculated absolutely exactly (without using infinite summation in function h_1 or its approximation (38)) by solving the initial value problem (6) for the collisionless kinetic equation. This

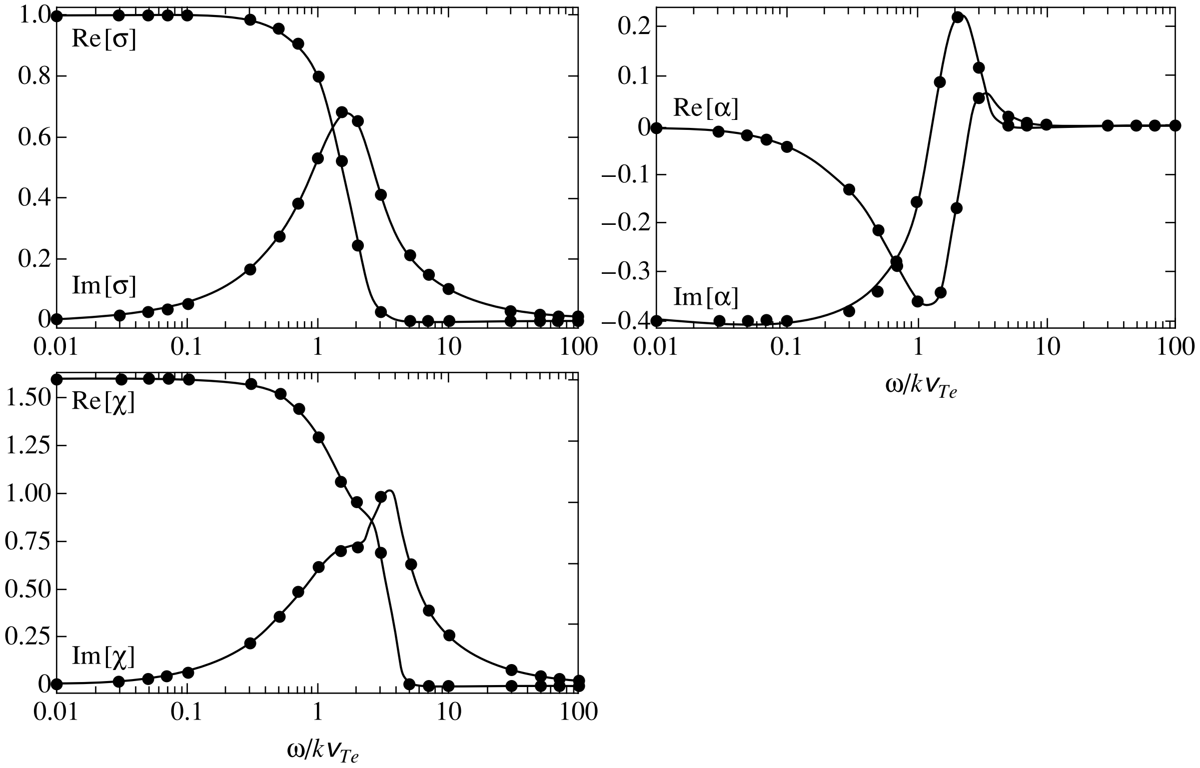


Fig. 3. Dependence of the real and imaginary parts of transport coefficients σ, α, and χ on ω/kv_{Te}, calculated using formula (38) (bullets), in comparison with the exact collisionless theory (40) (solid curve).

follows from the corresponding collisionless expressions for moments J_A^B,

$$J_N^N = \frac{i}{\omega}J_+(p), \quad J_N^T = \frac{i}{3\omega}((p^2-1)J_+(p)-p^2),$$
$$J_T^T = \frac{i}{9\omega}((p^4-2p^2+5)J_+(p)-p^4+p^2), \tag{40}$$

where

$$J_+(x) = x\exp\left(-\frac{x^2}{2}\right)\int_{i\infty}^{x} dt\exp\frac{t^2}{2}$$

is the standard dispersion function emerging in the collisionless theory of plasmas [1]. The behavior of collisionless transport coefficients as functions of ω/kv_{Te} is illustrated in Fig. 3. These coefficients correspond to the results obtained in [20]. In the quasi-static collisionless limit, we represent the result

$$\sigma = \frac{5}{\sqrt{8\pi}}\frac{e^2 n_e v_{Te}}{kT_e}, \quad \alpha = -\frac{1}{\sqrt{2\pi}}\frac{en_e v_{Te}}{kT_e},$$
$$\chi = \frac{4}{\sqrt{2\pi}}\frac{n_e v_{Te}}{k} \tag{41}$$

following from (40) [16, 19] and corresponding to the free (Knudsen) quasi-stationary transport.

The expression for the heat flux is often written in terms of the temperature gradient and electric current [14]. Accordingly, eliminating the electric field from Eqs. (30), we obtain

$$\mathbf{q} = -\frac{\alpha T_e}{\sigma}\mathbf{j} - \kappa i\mathbf{k}\delta T_e - n_e T_e \beta u_i,$$
$$\kappa = \chi - \frac{\alpha^2 T_e}{\sigma}, \quad \beta = \beta_q - \frac{e\alpha}{\sigma}\beta_j, \tag{42}$$

where the heat conductivity κ and the ion convective transport coefficient β are introduced. Figure 4 illustrates the dependence of these quantities on ω for two different collision parameters. Formulas (42) for $j = 0$ (no-current plasma) are directly related to the description of transport in a ICF plasma. It was shown in [17] using the heat conductivity in the problem of thermal relaxation as an example that transient effects are important for $k\lambda_{ei} \gtrsim 0.1$. For such inhomogeneity scales, the quasi-stationary approaches that have mainly been developed until now [9, 11, 12, 15, 16] are inapplicable. The theory developed above suspends this constraint. The equations of nonlocal hydrodynamics with nonstationary transport coefficients proposed by

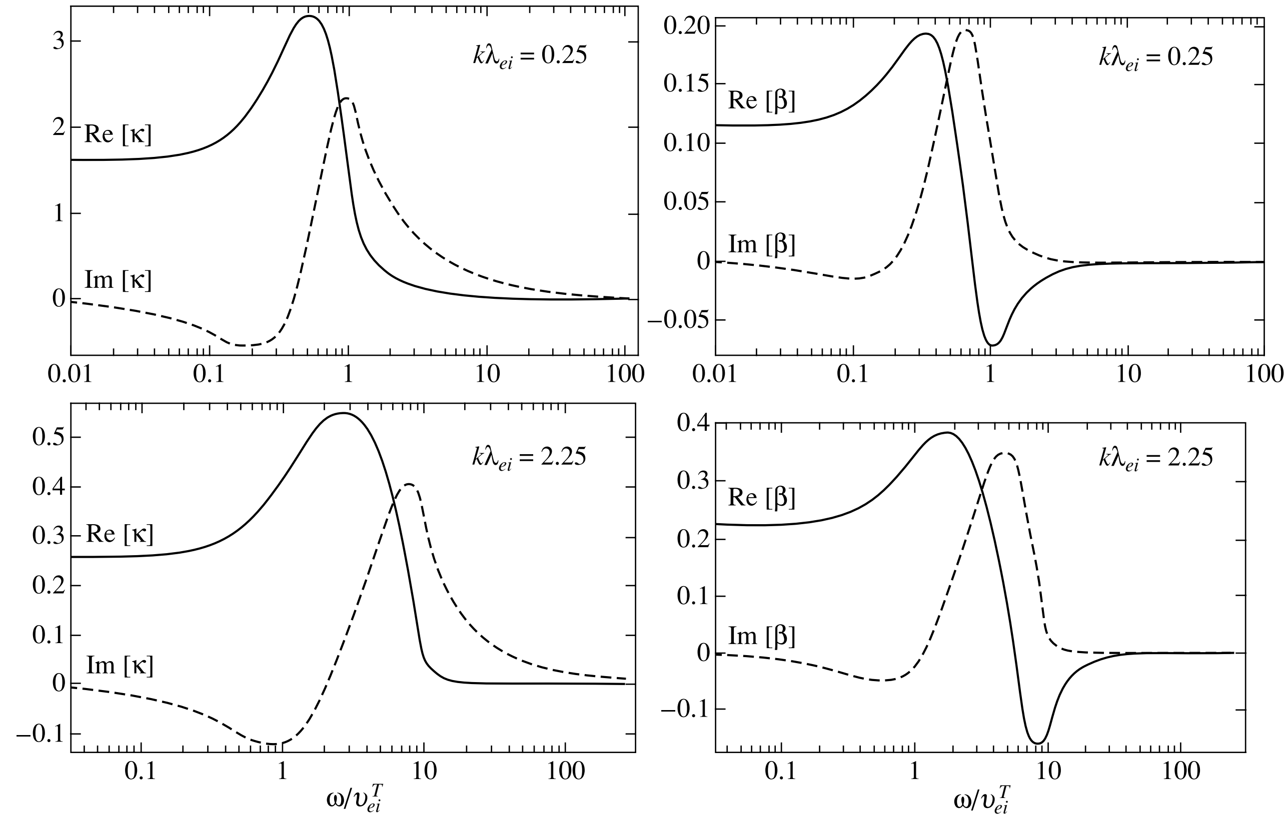


Fig. 4. Dependence of the real and imaginary parts of transport coefficients κ and β on ω/ν_{ei}^T for $k\lambda_{ei} = 0.25$ and $k\lambda_{ei} = 2.25$.

us make it possible to describe a plasma for any spatial and temporal perturbation scales.

4. LONGITUDINAL PERMITTIVITY OF PLASMAS

Since the hydrodynamic equations (28) are equivalent to a kinetic description and completely determine the linear response of a plasma to small perturbations in the entire range of parameters (ω, k), these equations can be used for obtaining the longitudinal permittivity $\epsilon(\omega, k)$ of the plasma. To calculate the permittivity

$$\epsilon = 1 + 4\pi i \frac{j}{\omega E}, \tag{43}$$

we eliminate the density and electron temperature perturbations from the expression for current, solving system (28):

$$\begin{aligned} j = {} & \left[1 - i\omega\left(\frac{e^2 n_e}{k^2 T_e \sigma} + \frac{2n_e(\sigma + e\alpha)^2}{\sigma^2(2k^2\kappa - 3i\omega n_e)}\right)\right]^{-1} \\ & \times \left\{ -\frac{ie^2 n_e}{k^2 T_e}\omega E \right. \\ & \left. + e n_e u_i \left[1 - i\omega\left(\frac{e^2 n_e \beta_j}{k^2 T_e \sigma} + \frac{2n_e(\sigma + e\alpha)(1 - \beta)}{\sigma(2k^2\kappa - 3i\omega n)}\right)\right]\right\} \\ \equiv {} & -\frac{ie^2 n_e}{k^2 T_e}\omega(1 + i\omega J_N^N)E + e n_e u_i(1 + i\omega J_N^R). \end{aligned} \tag{44}$$

Let us first analyze a purely electron plasma in the limit of stationary (infinitely heavy) ions, when $u_i = 0$.

A. Electronic Permittivity

We will characterize the partial electron contribution $\delta\epsilon_e$ to permittivity ($\epsilon = 1 + \delta\epsilon_e$) by quantity $\delta\epsilon = k^2\lambda_{De}^2\delta\epsilon_e$, where λ_{De} is the Debye radius for electrons. Using relation (44), we obtain for this quantity the expression

$$\begin{aligned} \delta\epsilon = {} & \left[1 - i\omega\left(\frac{e^2 n_e}{k^2 T_e \sigma} + \frac{2n_e(\sigma + e\alpha)^2}{\sigma^2(2k^2\kappa - 3i\omega n_e)}\right)\right]^{-1} \\ & \equiv 1 + i\omega J_N^N, \end{aligned} \tag{45}$$

which makes it possible to find the contribution from all transport coefficients and is in fact determined by only one momentum J_N^N.

In the classical hydrodynamic limit $k\lambda_{ei} < 0.06/\sqrt{Z}$, $0.1\omega/\nu_{ei}^T$, the analytic expression for $\delta\epsilon$ is obtained by

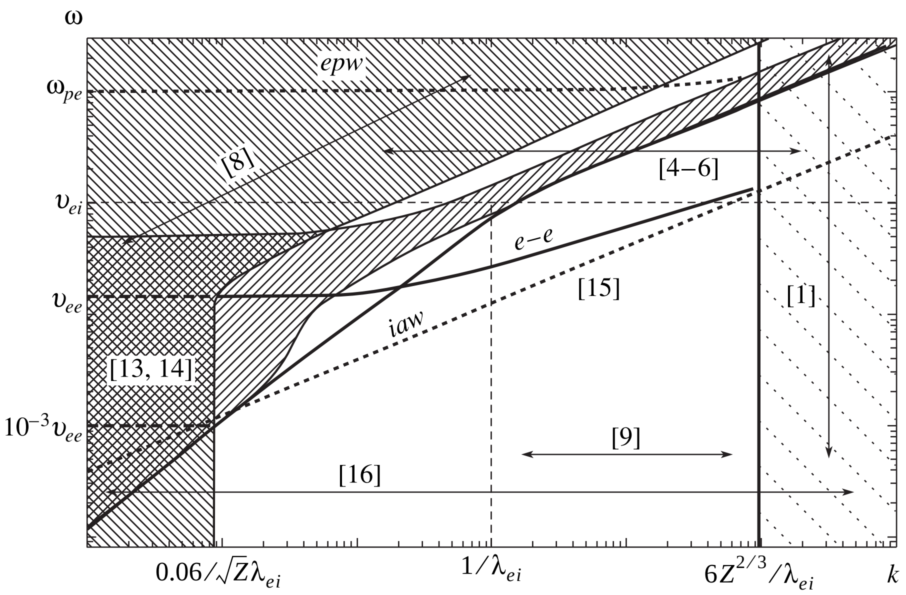


Fig. 5. Parametric (k, ω) plane for the longitudinal permittivity of plasmas. Dotted curves describe the spectra corresponding to the Langmuir (*epw*) and ion-acoustic waves (*iaw*). References are given in the brackets.

substituting Eqs. (34) into formula (45). In the limit of low frequencies ($\omega \ll \nu_{ei}^T$), this leads to the expression

$$\delta\epsilon = \frac{2x(8x-3i\omega)}{16x^2-6\omega^2-47i\omega x}, \quad x = \frac{32k^2v_{Te}^2}{3\pi\nu_{ei}^T}. \qquad (46)$$

In the limit $\omega \gg x$, the electron susceptibility is determined by the classical electrical conductivity $\delta\epsilon_e = 4\pi i\sigma_{SH}/\omega$. In the opposite case ($\omega \ll x$), the transport coefficients are immaterial for describing the static permittivity corresponding to Debye screening, $\mathrm{Re}\epsilon = 1 + 1/k^2\lambda_{De}^2$; these coefficients determine only the small imaginary correction $\mathrm{Im}\epsilon = 41\omega/16x$, which receives a comparable contribution from coefficients σ, α, and χ. The dispersion relation $\epsilon = 0$ for $k\lambda_{De} \ll 1$ gives the classical entropy mode $\omega = 2ik^2\kappa_{SH}/3n_e$ with a heat conductivity defined by relation (42), $\kappa_{SH} = 128n_e v_{Te}\lambda_{ei}/3\pi$ [13, 14]. For fast processes ($\omega \gg \nu_{ei}^T$), the permittivity is determined by the high-frequency electrical conductivity and is described by the well-known expression $\epsilon = 1 - (\omega_{pe}^2/\omega^2)(1 - i\nu_{ei}^T/\omega)$ [7].

Figure 5 shows the parametric (k, ω) plane divided into regions corresponding to different approximations for describing the permittivity beginning from the classical hydrodynamic limit (left-hatched region) to the collisionless kinetic limit (dot-hatched region). The region between the fine solid curves in Fig. 5 (right hatching) corresponds to strongly decaying perturbations, for which $\mathrm{Im}\epsilon > \mathrm{Re}\epsilon$. Under the unmarked bold solid curve, the real part of the permittivity corresponds to the Debye screening, $\mathrm{Re}\epsilon = 1 + 1/k^2\lambda_{De}^2$. The $\omega(k)$ boundary curve denoted by *e–e* separates the quasistationary regime (33), for which electron–electron collisions are important, from the nonstationary regime. It should be noted that for $k\lambda_{ei} \ll 6Z^{2/3}$ in the quasi-static approximation, it is sufficient to use two angular harmonics (diffusion approximation) for calculating the electron distribution function and, accordingly, for all transport coefficients as well as permittivity [9]. In this limit and for $k\lambda_{ei} \gg 1/\sqrt{Z}$, the approximate expression for the permittivity has the form [9]

$$\delta\epsilon = 1 + i\frac{\omega}{kv_{Te}}\left\{\sqrt{\frac{\pi}{2}} + 2.17\frac{Z^{2/7}}{(k\lambda_{ei})^{3/7}}\right\}, \qquad (47)$$

which is close to the exact solution. Our analysis proves that the range of applicability of relation (47) is in fact slightly narrower, $k\lambda_{ei} > 1$.

In the frequency range in which the *e–e* collisions can be neglected, we can reconstruct from relation (37) the permittivity obtained in [4–6], which leads to the well-known expression

$$\epsilon = 1 + \frac{1}{k^2\lambda_{De}^2}\left[1 - J_+\left(\frac{\omega}{kv_{Te}}\right)\right] \qquad (48)$$

in the collisionless kinetic limit $\nu_{ei} \longrightarrow 0$.

The general expression derived for permittivity is applicable for describing the plasma in the entire range of k and ω for any number of collisions in the plasma also. The contribution of collisions to the permittivity of the plasma is still described in most cases by using a

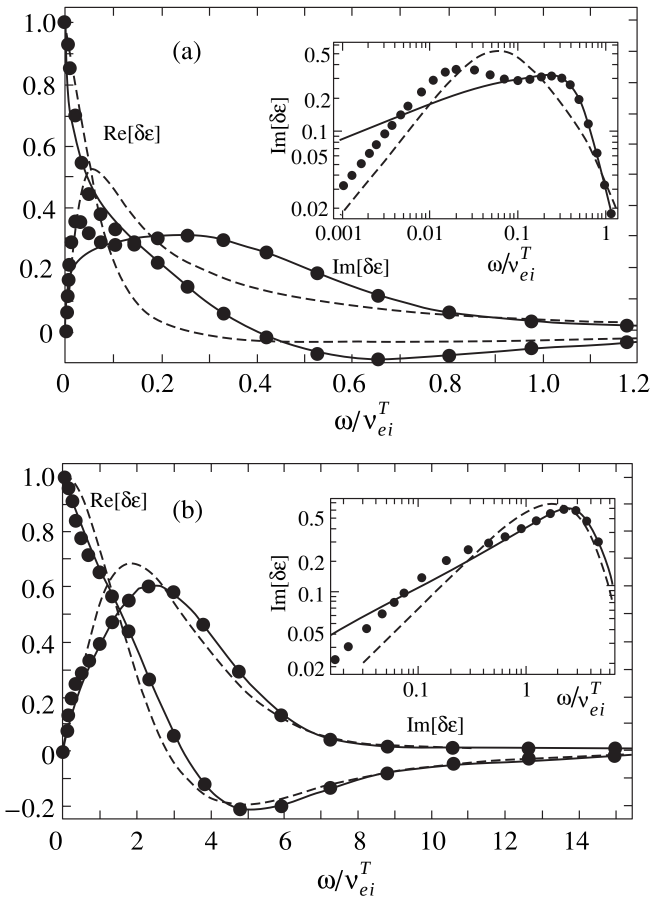


Fig. 6. Dependence of the real and imaginary parts of $\delta\epsilon(\omega, k)$ (45) (bullets) on ω/ν_{ei}^T for $k\lambda_{ei} = 0.25$ (a) and 2.25 (b) in comparison with the theory disregarding the electron–electron collisions (solid curves) [4] and the BGK model (49) (dashed curves) [3].

simplified BGK collision integral. The theory developed here makes it possible to determine the accuracy of this approximation. The best agreement is obtained using the expression proposed in [3]:

$$\delta\epsilon^C = \frac{1 - J_+(y)}{1 - iJ_+(y)/yk\lambda_{ei}}, \quad y = \frac{\omega + i\nu_{ei}^T}{kv_{Te}}. \tag{49}$$

However, in spite of the fact that the behavior described by this expression is qualitatively correct on the whole, it noticeably differs from the exact result for $k\lambda_{ei} < 1$ in a wide frequency range (see Fig. 6a). With increasing $k\lambda_{ei}$, the agreement is improved; however, it follows from Fig. 6b that the empirical formula (49) still differs from the exact solution by a factor of 2 to 3 in the range of parameters $\omega \sim \nu_{ei}$.

B. Ion Contribution to Permittivity

In accordance with definition (43), elimination of the ion velocity from expression (44) makes it possible to obtain the ion contribution to permittivity. Strictly speaking, a complete kinetic description of the ion component is required in this case. However, according to the results obtained in [25], for describing rapid perturbations ($\omega \gg kv_{Ti}$), where v_{Ti} is the ion thermal velocity, we can use the hydrodynamic equations for ions, in which the ion viscosity and heat conductivity are taken into account using the 21-moment approximation of the Grad method. This makes it possible to write conservation laws (8), (9), and (12) for ions in the form

$$\begin{aligned}
-i\omega u_i &= \frac{eZ}{m_i}E - ikv_{Ti}^2\left(\frac{\delta n_i}{n_i} + \frac{\delta T_i}{T_i}\right) \\
&\quad - \frac{4}{3}\frac{k^2 v_{Ti}^2}{\nu_i}\hat{\eta}_i u_i + \frac{1}{n_i m_i}\mathcal{R}_{ie} - \nu_{ei}^T u_i, \\
-i\omega\delta T_i &= -\frac{2}{3}iku_iT_i - \frac{2}{3n_i}k^2\kappa_i\delta T_i, \\
-i\omega\delta n_i + iku_in_i &= 0.
\end{aligned} \tag{50}$$

The contribution from collisions to the momentum conservation law (12) is described by Eq. (18); the longitudinal component of the stress tensor is represented in terms of the ion viscosity $\Pi_\parallel^i = ik4n_iT_i\hat{\eta}_iu_i/3\nu_i$ [25],

$$\hat{\eta}_i = \frac{i\nu_i(\omega + 1.46i\nu_i)}{(\omega + 1.20i\nu_i)(\omega + 1.46i\nu_i) + 0.23\nu_i^2}; \tag{51}$$

the energy exchange during the *e–i* collisions in the energy conservation law (9) is assumed to be negligibly weak; and the expression for the ion heat flux $q_i = -ik\kappa_i$ is determined by the heat conductivity [25]

$$\kappa_i = \frac{5}{2}\frac{n_iv_{Ti}^2}{\nu_i}\frac{i(\omega + 1.29i\nu_i)}{(\omega + 0.8i\nu_i)(\omega + 1.29i\nu_i) + 0.21\nu_i^2}. \tag{52}$$

Here, the ion–ion collision frequency is introduced in the standard manner, $\nu_i = 4\sqrt{\pi}e_i^4n_i\Lambda/3T_i^{3/2}\sqrt{m_i}$.

Using relations (50)–(52), we can describe the total permittivity of the plasma in the form

$$\epsilon = 1 + \frac{1 + i\omega J_N^N}{k^2\lambda_{De}^2} - \frac{c_s^2}{\lambda_{De}^2}\frac{(1 + i\omega J_N^R)^2}{\Delta}, \tag{53}$$

where

$$\Delta = \omega^2 + i\omega\nu_{ei}^T + i\omega k^2 c_s^2\left(J_R^R + \frac{(2\pi)^{3/2}}{n_e k^2 \lambda_{ei}}\int_0^\infty dv\,v\frac{\nu_{ei}}{\nu_1}f_M^e\right) - k^2 v_{Ti}^2\left(\frac{4i\omega\hat{\eta}_i}{3\nu_i} - \frac{5n_i\omega + 2ik^2\kappa_i}{3n_i\omega + 2ik^2\kappa_i}\right). \quad (54)$$

Thus, expression (53) defines the total longitudinal permittivity of a plasma with negligibly small Landau damping at ions, $\omega \gg kv_{Ti}$.

5. CONCLUSIONS

Using the kinetic approach, we have formulated the equations of nonlocal transport for small perturbations in the general case of arbitrary relations between the characteristic space and time scales of plasma inhomogeneity and particle collisions. Consequently, the derived quasi-hydrodynamic equations are in fact equivalent to the description of a plasma with the help of the kinetic equation since these equations consistently take into account the kinetic transport effects under the conditions and spatial and temporal nonlocality. The corresponding nonstationary and nonlocal transport coefficients in the Fourier representation are investigated in detail in the entire (ω, k) region with visual graphic illustrations. The developed theory generalizes particular transport models constructed earlier to the case of arbitrary (ω, k) and describes all limiting transitions to the known results.

We propose a convenient algorithm for calculating the longitudinal permittivity of a plasma for arbitrary values of frequency and wavenumber. The expression for permittivity derived here describes a smooth transition from the hydrodynamic region of strong collisions to the collisionless kinetic region and from the static to the high-frequency limit. On the basis of our theory, it becomes possible to analyze the plasma response in an intermediate region of wavenumbers and frequencies, in which the values of these quantities are on the order of the inverse mean free path and electron collision frequency, respectively. A qualitative description of the plasma response is especially difficult for such parameters.

The permittivity is determined on the basis of the Landau approximation for the collision integral, although a more exact Balesku–Lenard expression for this integral [26, 27] taking into account the effect of dynamic polarization of the plasma is itself, strictly speaking, a function of permittivity. The solution of the problem is complicated under the conditions when this effect is significant. This complication is associated with allowance for the contribution from the interaction of particles with plasma oscillations. Since the phase velocity of Langmuir oscillations of electrons is larger than the electron thermal velocity, the number of particles interacting with plasma waves is relatively small and the dynamic polarization effect of the plasma does not noticeably affect the collisions. This effect might become significant in the case of interaction with slow ion–acoustic waves [28]. The corresponding correction to the Landau collision integral is most substantial for the electron–electron collision integral and becomes predominant for a strongly nonisothermal plasma, when $ZT_e/T_i > \Lambda_{ee}\ln(Z^2 m_i T_e^3/m_e T_i^3)$ (i.e., for $ZT_e/T_i > 10^3$) [28]. Thus, the theory developed here is applicable when $ZT_e/T_i < 10^3$.

The development of nonlocal hydrodynamics is especially important for describing heat transport playing a decisive role in the ICF problem (in particular, for interpreting and simulating experiments with laser plasmas). It is well known that traditional hydrodynamic codes with a heat flux, which can be described on the basis of the classical theory or its simple heuristic modifications, fail to explain experimental data correctly, while the model with nonlocal transport give a much better agreement with experiment [29]. This raises hopes that the nonstationary transport theory presented here has a high potential for practical application in the ICF problem. Analysis of transport in the case of small-scale plasma inhomogeneities must occupy a special place in this case. A closely related problem is that of energy transport in a speckle laser beam used for obtaining a more uniform irradiation of a thermonuclear target with the help of a number of laser radiation smoothing mechanisms. It has been demonstrated that the description of temperature relaxation in small-scale $(L \sim \lambda_{ei})$ hot laser spots requires taking into account the effects associated with nonstationary nature of transport coefficients [17]. These effects are important for describing rapid heating of a hot laser plasma, including the heating of the skin layer of the plasma with a sharp boundary (the latter is characteristic of the interaction of short laser pulses with solid targets).

Direct application of nonlocal nonstationary transport coefficients for small perturbations may involve the development of a theory of laser plasma instabilities. The importance of allowance for nonlocality of transport in the quasi-stationary limit of filamentation instability and the stimulated Brillouin scattering was demonstrated in [30, 31]. The approach worked out by us paves the way for the development of such a theory for a strongly nonstationary laser plasma. Another important trend in the application of permittivity is associated with calculation of the Thomson scattering cross section [32], which is widely used for diagnostics of plasmas [29].

ACKNOWLEDGMENTS

This study was financed by the Russian Foundation for Basic Research (project no. 03-02-16428) and the Natural Science and Engineering Research Council of Canada (NSERC).

REFERENCES

1. A. F. Aleksandrov, L. S. Bogdankevich, and A. A. Rukhadze, *Principles of Plasma Electrodynamics*, 2nd ed. (Vysshaya Shkola, Moscow, 1988; Springer, Berlin, 1984).
2. E. M. Lifshitz and L. P. Pitaevskiĭ, *Physical Kinetics* (Nauka, Moscow, 1979; Pergamon, Oxford, 1981).
3. M. Opher, G. J. Morales, and J. N. Leboeuf, Phys. Rev. E **66**, 016407 (2002).
4. V. Yu. Bychenkov, Fiz. Plazmy (Moscow) **24**, 862 (1998) [Plasma Phys. Rep. **24**, 801 (1998)].
5. R. A. Koch and W. Horton, Jr., Phys. Fluids **18**, 861 (1975).
6. J. R. Penano, G. J. Morales, and J. E. Maggs, Phys. Plasmas **4**, 555 (1997).
7. V. P. Silin and A. A. Rukhadze, *Electromagnetic Properties of Plasma and Plasma-like Media* (Atomizdat, Moscow, 1961), p. 110 [in Russian].
8. I. P. Shkarofsky, T. W. Johnston, and M. P. Bachynski, *The Particle Kinetics of Plasmas* (Addison-Wesley, Reading, 1966).
9. V. P. Silin, Usp. Fiz. Nauk **172**, 1021 (2002) [Phys. Usp. **45**, 955 (2002)].
10. R. Bell, Phys. Fluids **26**, 279 (1983).
11. J. F. Luciani, P. Mora, and J. Virmont, Phys. Rev. Lett. **51**, 1664 (1983).
12. J. R. Albritton, E. A. Williams, I. B. Bernstein, and K. P. Swartz, Phys. Rev. Lett. **57**, 1887 (1986).
13. L. Spitzer and R. Härm, Phys. Rev. **89**, 977 (1953).
14. S. I. Braginskiĭ, in *Problems of the Plasma Theory* (Atomizdat, Moscow, 1963), Vol. 1, p. 183 [in Russian].
15. E. M. Epperlein, Phys. Plasmas **1**, 109 (1994).
16. V. Yu. Bychenkov, W. Rozmus, V. T. Tikhonchuk, and A. V. Brantov, Phys. Rev. Lett. **75**, 4405 (1995); A. V. Brantov, V. Yu. Bychenkov, V. T. Tikhonchuk, and W. Rozmus, Zh. Éksp. Teor. Fiz. **110**, 1301 (1996) [JETP **83**, 716 (1996)]; V. Yu. Bychenkov, V. N. Novikov, and V. T. Tikhonchuk, Zh. Éksp. Teor. Fiz. **114**, 1691 (1998) [JETP **87**, 916 (1998)].
17. S. Brunner, E. Valeo, and J. A. Krommes, Phys. Plasmas **7**, 2810 (2000).
18. A. Bendib, G. Matthieussent, and F. Bouzid, Phys. Plasmas **9**, 35 (2002).
19. G. W. Hammett and F. W. Perkins, Phys. Rev. Lett. **64**, 3019 (1990); G. W. Hammett, W. Dorland, and F. W. Perkins, Phys. Fluids B **4**, 2052 (1992).
20. Z. Chang and J. D. Callen, Phys. Fluids B **4**, 1167 (1992).
21. N. Mattor and S. E. Parker, Phys. Rev. Lett. **79**, 3419 (1997).
22. R. D. Hazeltine, Phys. Plasmas **5**, 3282 (1998).
23. V. Yu. Bychenkov, V. T. Tikhonchuk, and W. Rozmus, Phys. Plasmas **4**, 4205 (1997).
24. J. F. Myatt, W. Rozmus, V. Yu. Bychenkov, and V. T. Tikhonchuk, Phys. Rev. E **52**, 6759 (1995).
25. V. Yu. Bychenkov, J. Myatt, W. Rozmus, and V. T. Tikhonchuk, Phys. Plasmas **1**, 2419 (1994).
26. R. Balesku, Phys. Fluids **3**, 52 (1960).
27. A. Lenard, Ann. Phys. (N.Y.) **3**, 90 (1960).
28. V. P. Silin, *Introduction to the Kinetic Theory of Gases* (Nauka, Moscow, 1971), p. 240 [in Russian].
29. G. Gregory, S. H. Glenser, J. Knight, *et al.*, Phys. Rev. Lett. **92**, 205006 (2004).
30. E. M. Epperlein, Phys. Rev. Lett. **65**, 2145 (1990).
31. V. Yu. Bychenkov, W. Rozmus, A. V. Brantov, and V. T. Tikhonchuk, Phys. Plasmas **7**, 1511 (2000).
32. J. F. Myatt, W. Rozmus, V. Yu. Bychenkov, and V. T. Tikhonchuk, Phys. Rev. E **57**, 3383 (1998).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 100, No. 6, 2005, pp. 1175–1187.
Translated from Zhurnal Éksperimental’noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 6, 2005, pp. 1334–1349.
Original Russian Text Copyright © 2005 by Lipson, Rusetskiĭ, Karabut, Miley.

STATISTICAL, NONLINEAR, AND SOFT MATTER PHYSICS

DD Reaction Enhancement and X-ray Generation in a High-Current Pulsed Glow Discharge in Deuterium with Titanium Cathode at 0.8–2.45 kV

A. G. Lipson[a,b], A. S. Rusetskiĭ[c], A. B. Karabut[d], and G. Miley[a]
[a]*Department of Nuclear, Plasma, and Radiological Engineering, University of Illinois at Urbana–Champaign, Urbana (IL) 61801, USA*
[b]*Institute of Physical Chemistry, Russian Academy of Sciences, Moscow, 119915 Russia*
[c]*Lebedev Physical Institute, Russian Academy of Sciences, Moscow, 119991 Russia*
[d]*“Luch” Research and Production Corporation, Podolsk, Moscow oblast, 142100 Russia*
e-mail: lipson@uiuc.edu

Received December 10, 2004

Abstract—The DD reaction yield (3-MeV protons) and the soft X-ray emission from a titanium (Ti) cathode surface in a periodic pulsed glow discharge in deuterium were studied at a discharge voltage of 0.8–2.45 kV and a discharge current density of 300–600 mA/cm^2. The electron screening potential U_e = 610 ± 150 eV was estimated in the range of deuteron energies 0.8 keV < E_d < 2.45 keV from an analysis of the DD reaction yield as a function of the accelerating voltage. The obtained data show evidence for a significant enhancement of the DD reaction yield in Ti in comparison to both theoretical estimates (based on the extrapolation of the known DD reaction cross section for $E_d \geq 5$ keV to low deuteron energies in the Bosch–Halle approximation) and the results of experiments using accelerators at the deuteron energies $E_{lab} \geq 2.5$ keV and current densities 50–500 μA/cm^2. Intense emission of soft X-ray quanta (10^{13}–10^{14} s^{-1} cm^{-2}) was observed at an average energy of 1.2–1.5 keV. The X-ray emission intensity and the DD reaction yield enhancement strongly depend on the rate of deuterium diffusion in a thin subsurface layer of Ti cathode. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The results of recent experiments at relatively low energies of bombarding particles [1–10] showed evidence for significant screening effects in metal targets, in particular, in the case of deuteron bombardment. Investigations of the fusion of light nuclei in metal targets showed exponential growth of the DD reaction yield enhancement (astrophysical S factor) with decreasing particle energy. Even in gaseous deuterium targets, the screening potential U_e = 25 ± 5 eV calculated for a deuteron projectile energy that has decreased down to E_{lab} = 3 keV is significantly higher than the adiabatic limit of the DD reaction in deuterium molecule (U_{ad} = 14.0 eV) [1].

In most metals and in some oxides [6–10], the experimental DD reaction yield enhancement and the screening potential are significantly higher than the values measured for gas targets and those theoretically predicted using extrapolation of the reaction cross section to low deuteron energies [10]. For example, in a PdO target bombarded with deuterons at an energy of $E_{lab} \geq 2.5$ keV, the experimental yield of the D(d, p)T reaction corresponds to a screening potential of U_e = 600 eV [8]. This leads to a 50-fold DD reaction yield enhancement as compared to that predicted using a standard cross section for this reaction calculated via extrapolation of the known cross section for $E_d \geq 5$ keV to low deuteron energies (E_{lab} = 2.5 keV) in the Bosch–Halle approximation [11].

Raiola *et al.* [6–8] systematically studied the DD reaction yields and the screening potentials deduced from the S factor for more than 40 elements of the Periodic table, including various metals and nonmetals. It was established that most of the studied metals possess a high screening potential ($U_e \geq 100$ eV) except for the metals of groups IV (Ti, Zr, Hf) and XI (Cu, Ag, Au) and some semiconductor and dielectric targets. No specific experimental conditions (including accelerator current density) and target properties (crystal and magnetic structure, charge number, and deuterium mobility in the target) were found to influence the DD reaction enhancement and the screening potential growth. It should be noted that the accelerator used in [6, 7] allowed the deuteron current only within 1–54 μA to be reached, so that the DD reaction yield could be measured only at a relatively high energy (E > 5 keV).

Kasagi *et al.* [9, 10] used a low-energy high-current accelerator (with a beam current within 60–400 μA) and measured the D(d, p)T reaction yields in some metals and oxides for $E_{lab} \geq 2.5$ keV. It was found that the screening potential value at these beam intensities strongly depends on the deuterium mobility in the target. In metals (Ti, Au) with low deuterium mobilities and high activation energies for deuteron diffusion, the screening potentials were low (U_e = 65 ± 15 and 70 ± 10 eV, respectively). These U_e values were only two times higher than the value for a gas (D_2) target. In contrast, the screening potentials in targets (such as Pd and PdO) with high deuterium mobilities were rather high (U_e = 310 and 600 eV, respectively) [10].

Unlike the aforementioned investigations [6–10] with metal targets and low-current accelerators, Bystritsky *et al.* [12–14] studied the D(d, n)He^3 reaction yield in a deuterated polyethylene (CD_n) target using the Z-pinch technique [12]. For the deuteron energies 3.6 keV < E_{lab} < 7.8 keV, the measured neutron yield and the evaluated *S* factor showed (to within the experimental error) the absence of enhancement of the DD reaction: the D(d, n)He^3 reaction cross section in CD_n was comparable to that described using the Bosch–Halle extrapolation in the given range of deuteron energies [14]. It should be noted that the dielectric target used in [13, 14] was characterized by a low mobility of deuterium, while the energy spread of bombarding deuterons was very large as compared to that in the experiments using accelerators [6–10].

Previously, accelerators with a low (<1%) spread of deuteron energies allowed the DD reaction yield in metal targets to be studied only for E_{lab} > 2.5 keV. Further decrease in the accelerating voltage leads to insurmountable difficulties in maintaining a sufficiently high beam current density, which makes impossible measurement of the DD reaction yield within an acceptable period of time because of an extremely low yield. At the same time, investigations of the DD reaction yield and cross section at low deuteron energies (below 1 keV) is of considerable interest from the standpoint of astrophysical processes of star evolution [15] and controlled thermonuclear reactions, in particular, the cross sections of a hot deuterium plasma interaction with a reactor wall [16].

An alternative possibility for studying the DD reaction yield at deuteron energies below 1 keV is offered by the experiments with a high-current pulsed discharge in deuterium. The results of previous experiments [17] showed that pulsed glow discharge makes it possible to obtain ions with the energies within 0.8–2.5 keV and current densities within 300–600 mA/cm^2 at a deuterium pressure of 1–10 Torr. The current density used fore the bombardment of the cathode (target) surface in glow discharge is three orders of magnitude higher than that accessible using accelerators. Preliminary estimates show that high-current bombardment of the cathode with deuterium ions in glow discharge can provide for detection of the DD reaction products even at $E \leq 1$ keV for exposures not exceeding several tens of hours (in the case of exponential growth of the DD reaction enhancement factor at low energies). Moreover, this bombardment may initiate the X-ray emission (accompanying the DD reaction initiation in a metal lattice with high solubility of hydrogen [18]), which has never been detected in the experiments using accelerators because of insufficiently high deuteron beam current density.

This paper presents the results of our systematic investigation of the DD reaction yield and X-ray emission in a titanium (Ti) cathode bombarded with deuterons of very low energies (0.8 keV < E_d < 2.45 keV) in a high-current pulsed glow discharge. The thick target yield (Ti cathode) showed evidence for a very high DD reaction enhancement described by screening potential U_e = 610 ± 150 eV. At E_d = 1.0 keV, the DD-reaction rate is nine orders of magnitude higher than that calculated by standard extrapolation [11] of the DD-reaction cross section.

2. EXPERIMENT

The yields of charged particles and X-ray quanta from a DD reaction were studied in a vacuum chamber where glow discharge was initiated at various voltages and currents. Figure 1 shows a schematic diagram of the experimental setup and the arrangement of detectors. The distance between the mobile Mo anode and the replaceable Ti cathode was 4–5 mm. The cathode was made of a 0.01-cm-thick cold rolled foil (99.95% Ti) and had an area of 0.64 cm^2). In order to eliminate overheating of the electrodes, thus reducing their sputtering rate and prolonging work life, the hollow cathode and anode holders were cooled from inside by a flow of distilled water. In order to prevent arc discharge formation at high current densities, the cathode and anode holders were covered by Teflon.

Rectangular voltage and current pulses with a short front (below 1 μs) and a duration of 200–400 μs were generated with a frequency of 3 kHz. The pulse parameters were monitored with the aid of a two-channel 100-MHz storage oscilloscope. A power supply source provided stable glow discharge at deuterium concentrations within 2–9 Torr. It was found that continuous leveling of the pressure during the glow discharge operation suppressed uncontrolled current and voltage fluctuations, thus stabilizing the discharge conditions. Under conditions of quasi-stable glow discharge, the average deuteron energy in the laboratory frame ($E_{lab} = eV$, where e is the electron charge) corresponds with high precision to the applied voltage V. Indeed, a low ($<10^{-5}$) degree of deuterium ionization in glow discharge makes the “maxwellization” effects (generating high-energy deuterons at the “tail” of the energy distribution) insignificant [19], so that the average deuteron energy is close to the nominal discharge voltage. The spread of the average deuteron energy, which is an analog of the particle energy spread in an accelerator) in our case did not exceed ±15% of the nominal value and

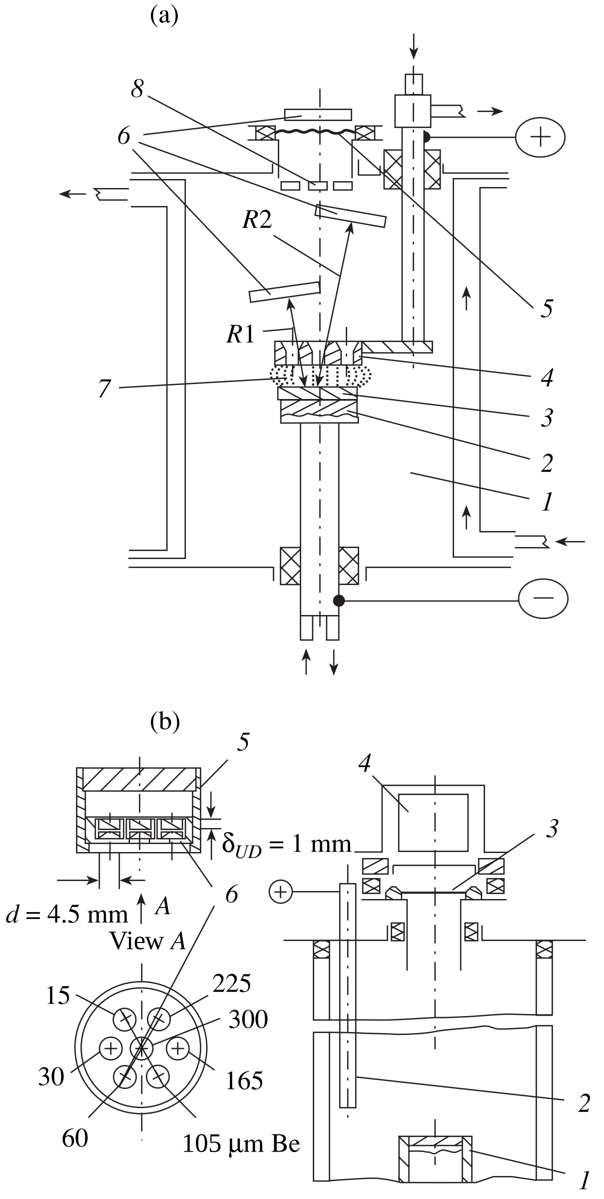


Fig. 1. (a) Schematic diagram of the glow discharge setup: (*1*) vacuum chamber; (*2*) cathode holder; (*3*) cathode; (*4*) anode; (*5*) Be window; (*6*) CR-39 detectors; (*7*) glow discharge region; (*8*) thermoluminescent detectors (TLDs) with 15- to 300-μm-thick Be filters. (b) Schematic diagram of the experiments with open cathode: (view *A*) TLDs with Be filters of different thickness; (*1*) cathode; (*2*) anode; (*3*) Be filters; (*4*) TLDs or pinhole camera; (*5*) metal holder of detectors; (*6*) 15- to 300-μm-thick Be filters.

was mostly determined by the residual pressure instability in the discharge chamber.

The current (I) and voltage (V) measurements in glow discharge at a constant pressure showed that the I–U curves were linear [19, 20] in the range of I = 100–300 mA and U = 800–2000 V (Fig. 2). The proportionality of current and voltage provides a convincing evidence for the absence of "arc" effects (capable of distorting the deuteron spectrum) in glow discharge. In order to provide for the maximum DD reaction yield, we selected the maximum current at a given pressure, so as to maintain a preset voltage according to the I–V curve.

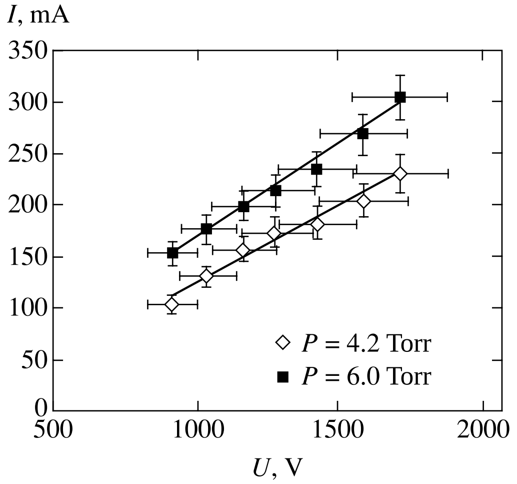


Fig. 2. Current-voltage curves of a glow discharge with Ti cathode in deuterium at a pressure of 4.2 and 6.0 Torr.

The results of temperature measurements performed using thermoresistors pressed against the rear side of the cathode showed that the cathode temperature was increased by 50–100 K over discharge operation [21]. However, these measurements do not provide correct estimates of the temperature of a near-surface layer of the cathode: examination of the cathode showed evidence of the material melting over the entire surface even at a minimum electric power (100 W/cm^2) applied to the discharge. In connection with this, below we will assume the maximum temperature in the near-surface layer of the cathode (with a thickness on the order of the range of bombarding deuterons in the cathode material) to be equal to the melting temperature of titanium.

In order to suppress spurious electromagnetic signals induced by the discharge, which are capable of significantly distorting the measured output signals, we did not use surface-barrier Si detectors (typically employed in experiments [6–10]). The DD reaction products were detected with the aid of plastic track detectors of the CR-39 type (Fukuvi Chemical Industry, Japan), which are insensitive to electromagnetic fields. These detectors were arranged in the discharge chamber behind the anode (in which seven holes were made) at a distance of 3 cm from the cathode surface (Fig. 1a). Measurements performed under analogous conditions of discharge in hydrogen (replacing deuterium) were used for determining the background level.

The CR-9 track detectors employed for detecting charged particles produced in the course of the D(d, p)T reaction were calibrated (Fig. 3) using standard α particle sources (E_α = 2.0–7.7 MeV), a cyclotron beam (E_α = 8.0–30.0 MeV), and a proton beam of the Van de Graaff accelerator ($E_{\rm p}$ = 0.5–3.0 MeV) of the

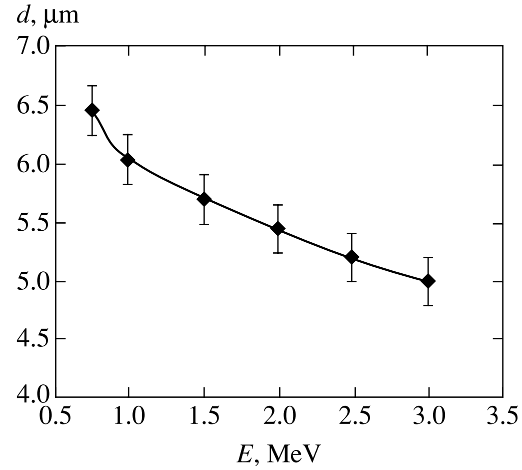


Fig. 3. Calibration curve of proton track diameter *d* versus particle energy *E* for CR-39 detectors.

Institute of Nuclear Physics (Moscow State University). After exposure, the detectors were etched in 6 M aqueous NaOH solution for 7 h at 70°C. The tracks ware examined and their diameters measured using an optical microscope equipped with a digital video camera.

In order to eliminate the action of discharge plasma and sputtered cathode particles on the detector surface, the detectors were screened with a 11-μm-thick Al foil. This foil transmits 3-MeV protons and absorbs 1-MeV tritons from the D(d, p)T reaction. The results of our measurements showed that the tracks due to 3-MeV protons transmitted through the foil have a diameter of about 5.2 μm. The detection efficiency of a CR-39 detector spaced by 3 cm from the cathode is determined primarily by the geometric factor. Taking into account the total area of holes in the anode, the geometric efficiency of 3-MeV proton detection was estimated at $\varepsilon_p = 5.6 \times 10^{-3}$.

The average energy and intensity of X-ray quanta emitted from the cathode surface were evaluated using Al_2O_3-based thermoluminescent detectors (TLDs) and a set of beryllium (Be) filters with thicknesses from 15 to 300 μm (2.8–55.5 mg/cm^2). These TLDs measured the absorbed radiation dose. Seven TLDs (each with a diameter of 5 mm) were arranged outside the discharge zone, at a distance of 7 cm from the anode. In a separate experiment performed in order to determine spatial position of the source of X-ray quanta in the discharge, the anode was shifted 20 mm away from the cathode and the TLD or a pinhole camera was positioned immediately in front of the cathode (Fig. 1b). The TLDs were calibrated using a standard ^{137}Cs source. The TLD signal readout and construction of the glow curves were performed using a special device based on a picosecond processor (Harshaw Co.).

The time correlation of X-ray emission and discharge current pulses were studied using a 17-mm-diam plastic (PMMA) scintillator and an FEU-85 photoelectron multiplier. These experiments were performed at a pressure of 4.2 Torr and a current of 250 mA. A positive image of the X-ray emitting zone was obtained with a pinhole camera using an X-ray-sensitive film.

The experiments devoted to the detection of charged particles were performed in a glow discharge operating at a voltage of 0.8–2.45 kV and a current of 240–450 mA. The duration of each exposure at a certain fixed discharge voltage was about 7 h. Preliminary experiments with CR-39 detectors covered by aluminum (Al) and polyethylene (PE) films of various thicknesses showed a statistically significant number of the 3-MeV proton tracks, which was dependent on the discharge voltage and current. Figure 4 presents typical distributions of the proton track diameters *d* in the detector covered by films of different thicknesses in a glow discharge in deuterium and hydrogen at $U = 1.25$ kV and $I = 240$ mA. As can be seen from these data, the track diameter strongly depends on the coating thickness, in accordance with the energy losses for 3-MeV protons [22] generated during operation of the glow discharge. In the presence of 11-μm-thick Al foil, the distribution peak is at $d = 5.2$ μm corresponding to $E_p = 2.85$ MeV (Fig. 4a). When the coating thickness was increased to 33 μm (Al foil) and 60 μm PE film, the peak shifts to $d = 6.8$ μm (Figs. 4b). In the glow discharge with Ti cathode in hydrogen under the same conditions (voltage, current, pressure) as in deuterium, no track were observed in the interval of diameters corresponding to the 3-MeV protons.

The thick target yield of 3-MeV protons $Y_t(E_d)$ from a Ti cathode bombarded by deuterons with an energy E_d was calculated using a formula [3]

$$Y_t(E_d) = \int_0^{E_d} N_D(x)\sigma_{lab}(E)\left(\frac{dE}{dx}\right)^{-1} dE, \qquad (1)$$

where $N_D(x)$, $\sigma_{lab}(E)$, and dE/dx are the deuteron concentration in the cathode, the DD reaction cross section, and the deuteron stopping power in titanium. The cross sections at low energies were determined by using the Bosch–Halle parametrization [11]. The deuteron stopping power in Ti target was assumed to be proportional to the particle velocity, which is consistent with the data available for various targets at low deuteron energies (down to $E_d = 1.0$ keV) [23, 24].

The yields of 3-MeV protons observed for various discharge voltages in the 0.8–2.45 kV range were normalized to the yield at the maximum voltage ($U = 2.45$ kV) with allowance for the discharge power and the effective temperature at the target surface (the factors influencing variations of the deuterium concentration $N_D(x)$ in the Ti cathode). The effective concentration of deuterium in Ti was defined as $N_D(\text{eff}) = k(W, T)N_D(x)$, where *T* and *W* are the temperature and power at the target surface, respectively. The coefficient $k(W, T)$ can be expressed as [21]

$$k(W, T) = \exp\left(-\frac{\varepsilon_d \Delta T}{k_B T_m T_0} \frac{W_m}{W_x}\right), \tag{2}$$

where k_B is the Boltzmann constant, $\varepsilon_d = 0.04$ eV is the activation energy for deuteron escape from the Ti cathode surface during discharge, $T_m = 1941$ K is the melting temperature of titanium, $T_0 = 290$ K is the initial temperature of the target, $\Delta T = T_m - T_0$, $W_m = 906.5$ W is the maximum discharge power at $E_d = E_m = 2.45$ keV, $I_m = 370$ mA, and W_x is the power at lower values of the current and voltage. The ε_d value was determined from data reported for the experiments using accelerators [9], by approximating the 3-MeV proton yield in Ti target with an Arrhenius plot [25] in the temperature interval 185–195 K at $E_d = 10$ keV (where no any DD enhancement takes place). The slope of the plot of yield versus temperature corresponds to the activation energy of the yield of deuterium from the target surface (Fig. 5).

3. EXPERIMENTAL RESULTS

The results of measurements using CR-39 detectors covered by 11-μm-thick Al foil showed a statistically significant number of 3-MeV proton tracks, which was dependent on the discharge voltage and current. Figure 6 presents a typical distribution of the track diameter for two discharge voltages ($U = 2175$ and 805 V) and the same current ($I = 250$ mA). A peak of the 3-MeV proton track diameter at $d = 5.2$ μm well agrees with the results of calibration (Fig. 2) and the preliminary measurements at $U = 1.25$ kV (Fig. 4).

The total set of data presented in Table 1 includes the numbers of tracks at various values of the current and voltage with allowance of the correction factor k calculated using formula (2). As the deuteron energy decreases from 2.45 to 0.8 keV, the 3-MeV proton yield drops by 3 orders in magnitude (with allowance for the normalization factor k). Calculated without the correction for k, the yield decreases by only one order in magnitude (Table 1, fourth column). This difference is related to the fact that the concentration of deuterium in Ti al low voltages (and discharge powers) is much higher than that at $U = 2.45$ kV because the effective temperature in the near-surface cathode layer is proportional to the discharge power.

Figure 7 shows the experimental yields of the D(d, p)T reaction in Ti as functions of the deuteron energy E_d in the range from 0.8 to 2.45 keV. Before the normalization using the coefficient k (Fig. 7a), the dependence of the 3-MeV proton yield on the discharge voltage has a more pronounced exponential character as compared to that expected taking into account the behavior of the cross section at low energies. After the

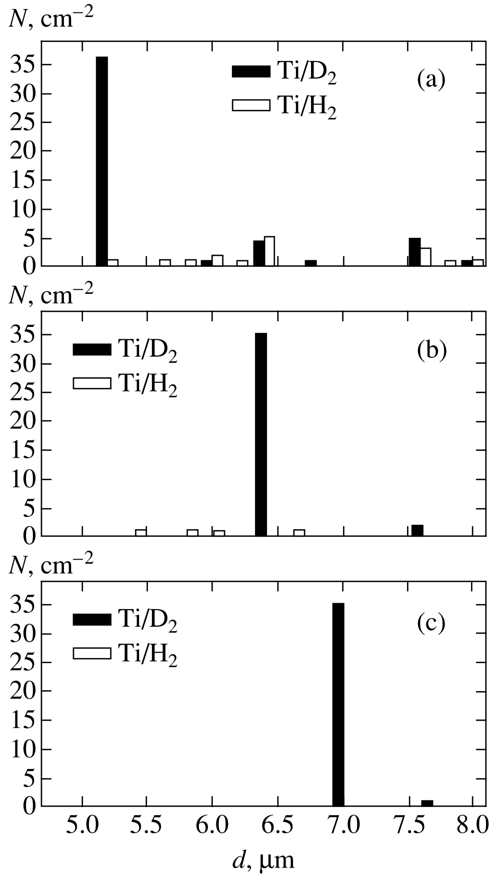


Fig. 4. Proton track diameter distributions in CR-39 detectors for glow discharge in deuterium and hydrogen (pressure, 6 Torr; $U = 1.25$ kV; $I = 240$ mA; exposure time, 7 h; cathode–detector distance, $R = 3$ cm): (a) detector covered with 11-μm-thick Al foil (the peak at $d = 5.2$ μm corresponds to protons with $E_p = 3$ MeV; energy losses in the foil are $\Delta E = 0.2 \pm 0.1$ MeV); (b) detector covered with 11-μm-thick Al foil and 60-μm-thick PE film (the peak at $d = 6.4$ μm corresponds to protons with $E_p = 3$ MeV; energy losses in the foil are $\Delta E = 1.1 \pm 0.2$ MeV); (c) detector covered with 33-μm-thick Al foil and 60-μm-thick PE film (the peak at $d = 6.8$ μm corresponds to protons with $E_p = 3$ MeV; energy losses in the foil are $\Delta E = 2.5 \pm 0.2$ MeV).

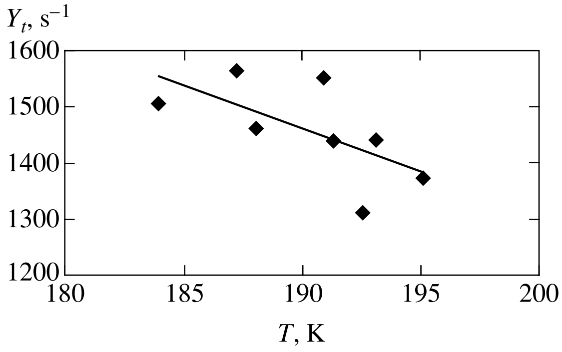


Fig. 5. Normalized proton yield versus temperature for a Ti target bombarded in an accelerator with deuterons at E_d = 10.0 keV and I = 60–100 μA [9]. Solid line corresponds to the Arrhenius function.

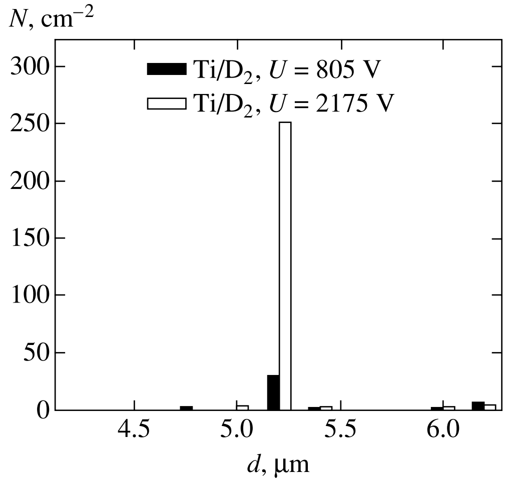


Fig. 6. Track diameter distribution for 3-MeV protons from a Ti cathode measured with a CR-39 detector for a discharge voltage of U_1= 805 and U_2 = 2175 V (I = 250 mA; exposure time, 7 h).

normalization (Fig. 7b), the curve $Y_t(E_d)$ curve exhibits a smoother behavior. This result confirms the need for normalization that takes into account the influence of random collisions (related to uncertainty in the deuterium concentration) leading to deviations of the experimental points from the exponent.

Figure 8 presents the DD proton yield Y_p at low deuteron energies normalized to that at 2.45 keV. For comparison, this figure also shows the standard yield of the D(d, p)T reaction (solid curve) calculated using the Bosch–Halle approximation [9, 10]. Even with allowance for the total error of measurements (including a systematic error and an error caused by the instability of glow discharge, amounting to ±10% in terms of the discharge voltage and current), the experimental plot of $Y_p/Y(2.45\text{ keV})$ lies well above the Bosch–Halle curve. This fact is clearly indicative of a large enhancement of the DD reaction in the near-surface of the target at low energies. In order to directly evaluate the DD reaction enhancement factor $f(E)$ and calculate the electron screening potential U_e in the 0.8–2.45 keV energy range, we used the formula [3]

$$f(E) = \frac{Y_p(E)}{Y_b(E)} = \exp\left[\pi\eta(E)\frac{U_e}{E}\right], \tag{3}$$

where $Y_p(E)$ is the experimental DD proton yield in a glow discharge, $Y_b(E)$ is the yield determined for the same energy using the Bosch–Halle extrapolation, $2\pi\eta = 31.29Z^2(\mu/E)^{1/2}$ is the Sommerfeld parameter, Z is the deuteron charge (for bombardment with D^+), and μ and E are the reduced mass and energy of deuteron, respectively.

Figure 9 shows the results obtained in an accelerator (curve *1*) [10] and in a glow discharge (curve 2) with the Ti cathode (target). In the case of measurements using the accelerator for 2.5 keV < E_d < 10.0 keV, the screening potential was $U_e = 65 \pm 10$ eV [9]. However, the screening potential evaluated from data on the reaction yield enhancement in glow discharge (Fig. 9, curve *2*) is $U_e = 610 \pm 150$ eV. For example, the experimental DD reaction yield enhancement at E_d = 1.0 keV is almost 9 orders of magnitude greater than that pre-

Table 1. The yield of 3-MeV protons from a Ti cathode for various voltages of glow discharge in deuterium

$\langle U\rangle$, V	$\langle I\rangle$, mA	W_m, W	N (5.2 μm), cm^2	$k(W, T)$	$\langle N_p\rangle$, s^{-1}	$[\langle n/\varepsilon\rangle \pm \sigma]$, s^{-1} (in 4π sr)	Y_p, C^{-1}
805	250	201.3	30	2.2×10^{-3}	2.6×10^{-6}	$(4.7 \pm 1.4) \times 10^{-4}$	1.9×10^{-3}
850	225	191.3	28	1.6×10^{-3}	1.8×10^{-6}	$(3.3 \pm 1.1) \times 10^{-4}$	1.5×10^{-3}
1000	370	370	35	3.6×10^{-2}	5.0×10^{-5}	$(9.0 \pm 1.9) \times 10^{-4}$	2.5×10^{-3}
1145	370	420	54	5.3×10^{-2}	1.1×10^{-4}	$(2.0 \pm 0.3) \times 10^{-2}$	5.3×10^{-2}
1190	240	286	30	1.3×10^{-2}	1.6×10^{-5}	$(3.0 \pm 0.5) \times 10^{-3}$	1.3×10^{-2}
1435	250	359	50	3.3×10^{-2}	7.0×10^{-5}	$(1.3 \pm 0.2) \times 10^{-2}$	5.2×10^{-2}
1500	450	675	71	0.16	4.5×10^{-4}	$(8.1 \pm 0.5) \times 10^{-2}$	1.8×10^{-1}
1647	300	495	62	8.3×10^{-2}	2.1×10^{-4}	$(4.0 \pm 0.5) \times 10^{-2}$	1.3×10^{-1}
2000	370	740	159	1.9×10^{-1}	1.2×10^{-3}	$(2.1 \pm 0.02) \times 10^{-1}$	5.7×10^{-1}
2175	250	544	252	1.1×10^{-1}	1.1×10^{-3}	$(2.0 \pm 0.02) \times 10^{-1}$	8.0×10^{-1}
2450	370	906.5	317	2.7×10^{-1}	3.4×10^{-3}	$(6.1 \pm 0.04) \times 10^{-1}$	1.65

Notes: $\langle U\rangle$, $\langle I\rangle$, and W_m are the average voltage, average current, and power of the glow discharge; N is the number density of 3-MeV proton tracks; $\langle N_p\rangle$ is the average count rate for 3-MeV protons; $\langle n/\varepsilon\rangle$ is the proton yield in a solid angle of 4π sr for a detection efficiency of $\varepsilon = 5{,}6 \times 10^{-3}$; Y_p is the DD proton yield per 1 C charge transferred by the deuteron current to the cathode.

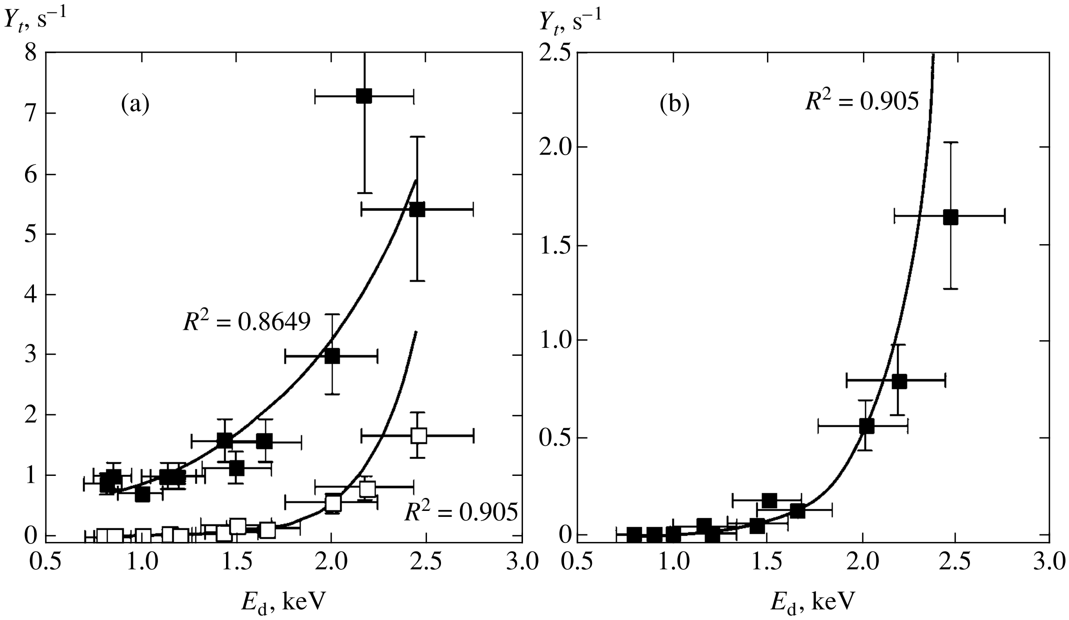


Fig. 7. The yield of 3-MeV protons from a Ti cathode: (a) nonnormalized; (b) normalized to deuterium concentration (with a normalization factor k determined using formula (2)).

dicted using the standard extrapolation of the reaction cross section to low deuteron energies.

Figure 10 presents the DD reaction yield in a glow discharge normalized to the yield measured in an accelerator at $E_d = 10.0$ keV (with correction of the DD proton yield for a lower effective target temperature in the accelerator as compared to that in glow discharge [9]). Similar to the situation in Fig. 8, the DD reaction yield in the glow discharge is much higher than that obtained by extrapolating to lower energy the value observed in the accelerator (for the screening potential $U_e = 65$ eV).

Thus, the data on the DD reaction yields in a glow discharge, corrected by normalization using the procedures analogous to those used in the experiments using accelerators [9, 10], demonstrate a much greater enhancement of the DD reaction yield at $E_d < 2.45$ keV as compared to that anticipated proceeding both from the theoretical extrapolation of the yield to low deu-

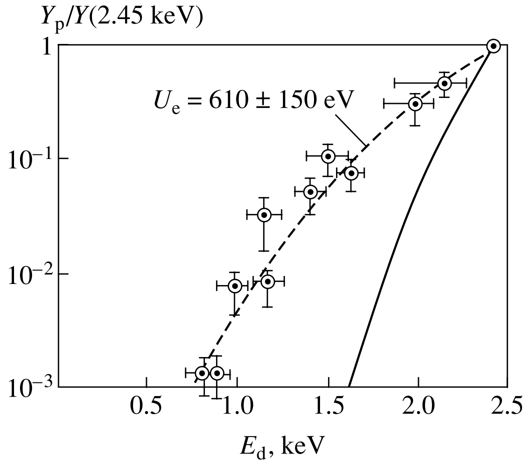


Fig. 8. The experimental yields of 3-MeV protons versus deuteron energy within 0.8 keV < E_d < 2.45 keV (normalized to the yield at $E_d = 2.45$ keV): Bosh–Halle approximation [11] (solid curve); DD reaction yield for a screening potential of $U_e = 610$ eV (dashed curve).

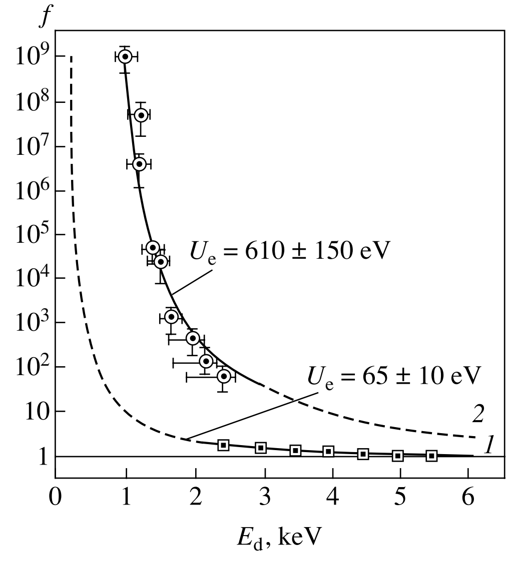


Fig. 9. Plot of the enhancement factor calculated using formula (3) versus E_d for a Ti target: (*1*) accelerator experiment [10]; (*2*) glow discharge. Solid curves correspond to the E_d intervals in which the yield was experimentally measured.

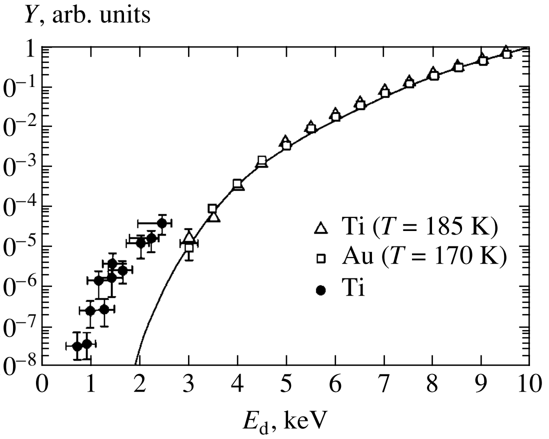


Fig. 10. Thick target yield of protons measured in glow discharge with Ti cathode (0.8 keV < E_d < 2.45 keV) and in accelerator experiments with Ti and Au targets (2.5 keV < E_d < 10 keV) [9]. All yields are normalized to the value obtained on the accelerator for E_d = 10 keV.

teron energies and from the analogous extrapolation of the yields obtained in the experiments using accelerators at E_d > 2.5 keV.

The experiments showed that the bombardment of a cathode in high-current periodic pulsed glow discharge is accompanied by the intense emission of soft X-ray quanta. In the experiments using TLDs in a glow discharge with a Ti cathode at U = 1.25 kV and I = 200 mA, we observed X-ray emission in the energy range of E_x = 1.1–1.4 keV and an intensity of I_x = 10^{13} s^{-1} in a solid angle of 4π (Fig. 11). It should be noted that the average energy of these radiation quanta is close to the energy of bombarding deuterons.

In order to determine the location of the X-ray source in the discharge, we used a setup in which the anode could be displaced relative to the cathode (Fig. 1b). In the case of a glow discharge with a "plasma" anode (i.e., with the anode shifted 20 mm away from the cathode) and the cathode open for monitoring with a pinhole camera, it was established that most X-ray quanta (>90%) are emitted predominantly from the cathode surface. A positive image of the open cathode obtained with X-ray-sensitive film (Fig. 12) represents a bright spot with dimensions corresponding to the diameter of the cathode spot.

The results of our experiments also showed that the X-ray emission pulses observed in a stationary discharge regime were strictly correlated with the current pulses. A growth in the discharge voltage and current was accompanied by substantially nonlinear growth in the yield of X-ray quanta (Fig. 13). The front of the signal from an X-ray radiation detector (based on a plastic scintillator and a photoelectron multiplier) usually coincided with the current pulse front. The X-ray intensity reaches maximum within several microseconds and then slowly decays over a period of about 200 μs.

It was found that the energy of the X-ray quanta, which was estimated using TLDs and a set of Be filters, exhibits a weak growth when the discharge voltage increases from 1.2 to 2.0 kV at a constant current of 200 mA. In order to provide for a change in the voltage at a constant current, the pressure of deuterium was varied in the interval from 2 to 9 Torr. It was found that, at U < 1.6 kV, the X-ray quantum energy E_x = 1.22 ± 0.15 keV is virtually independent of the discharge voltage. As the voltage is increased further, the X-ray quantum energy growth to reach E_x = 1.43 ± 0.17 keV.

The statement that the TLDs with Be filters detected X-ray quanta, rather than some other kind of ionizing radiation, can be confirmed by an analysis of the possible types of emission accompanying glow discharge. The TLDs employed are sensitive to X-ray and gamma

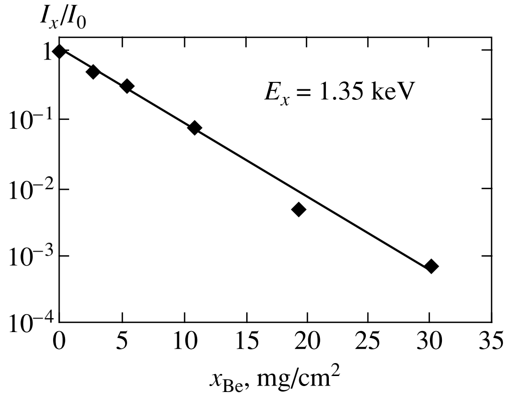


Fig. 11. Estimated energy E_x of X-ray quanta emitted from a Ti cathode in the glow discharge in deuterium (U = 1.25 kV; I = 200 mA; p = 4.4 Torr). The radiation was detected by a TLD with Be foils of variable thickness (see Fig. 1); I_x and I_0 are the X-ray intensity measured in front of and behind the filter.

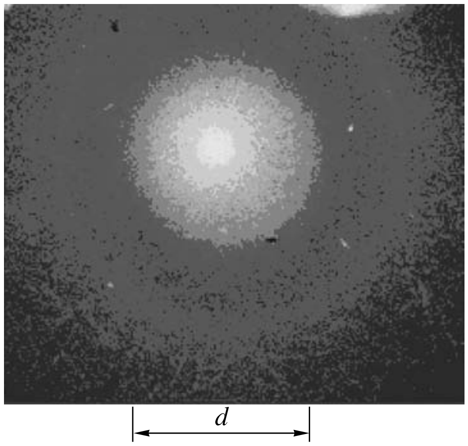


Fig. 12. A Ti cathode image obtained with X-ray film using 0.3-mm aperture pinhole camera covered by a 15-μm-thick Be foil (I = 150 mA; U = 1250 V; p = 5.3 Torr; exposure time, 1000 s).

quanta and to electrons with energies in the range from several units to several hundred of kiloelectronvolts. In the given interval of discharge voltages studied, it is unlikely that electrons can be accelerated to $E > 10$ keV. On the other hand, electrons with $E < 10$ keV are completely absorbed in a Be layer with a thickness of 15 μm. In our case, the intensity of radiation detected by TLDs exhibited a tenfold decrease only for a Be film thickness of $h > 100$ μm (Fig. 10). Therefore, we may ascertain that the radiation detected by TLDs represents only X-ray quanta, which are attenuated in Be films according to the well-known law.

The measurements of radiation doses I_x absorbed by TLDs for various discharge currents (100–270 mA) and voltages (1.0–1.8 kV) at a constant deuterium pressure of 6 and 4.2 Torr revealed exponential growth of I_x as a function of the effective discharge power $P^* = UIQ$, where Q is the on-off ratio for the discharge current pulses (Fig. 14). The yield of X-ray quanta at a constant pressure obeys the law

$$I_x = I_0\exp[(\varepsilon/k_{\mathrm{B}}T_{\mathrm{m}})P_x^*/P_0^*], \tag{4}$$

where $I_0 = 0.98$ and 0.725 Gy for $p = 6.0$ and 4.2 Torr, respectively; $\varepsilon = 0.04$ eV is the activation energy for deuteron escape from the Ti cathode surface (Fig. 5, formula (1)); $T_{\mathrm{m}} = 1941$ K is the melting temperature of titanium; and $P_0^* \approx 6.0$ W is the minimum (threshold) effective discharge power.

The efficiency of X-ray generation (the number of quanta emitted per implanted deuteron) as a function of the discharge current also obeys an exponential dependence with the same parameters ε, T_{m}, and P_0^* (Fig. 15). Thus, the X-ray yield strongly depends on the deuterium concentration at the target surface at the titanium melting temperature and exhibits a tendency to growth with increasing effective cathode temperature ($T_{\mathrm{eff}} \propto P^*$).

4. DISCUSSION OF RESULTS

The results of our experiments showed that a high-current glow discharge in deuterium at an applied voltage of 0.8–2.45 kV is characterized by significant enhancement of the DD reaction yield in a Ti cathode and is accompanied by intense X-ray emission. In contrast to the accelerator experiments [6–10], which were performed at much lower currents and higher deuteron energies and showed an enhancement of the DD reaction corresponding to an electron screening potential of $U_{\mathrm{e}} \le 65$ eV, our experiments with glow discharge gave a much higher value of $U_{\mathrm{e}} = 610 \pm 150$ eV in Ti.

Let us consider the possible sources of errors in the experiments with glow discharge, which might lead to overestimation of the screening potential and enhancement factor for the DD reaction in a Ti target. As was demonstrated in earlier accelerator studies [1, 3], the

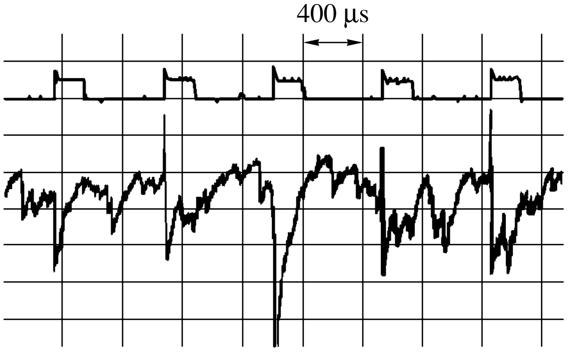


Fig. 13. Synchronized oscillograms of (*1*) X-ray emission pulses measured with the aid of a plastic scintillator and a photoelectron multiplier and (*2*) a glow discharge current ($\Delta\tau = 400$ μs) in deuterium ($U = 1.4$ kV; $I = 250$ mA; $p = 4.2$ Torr).

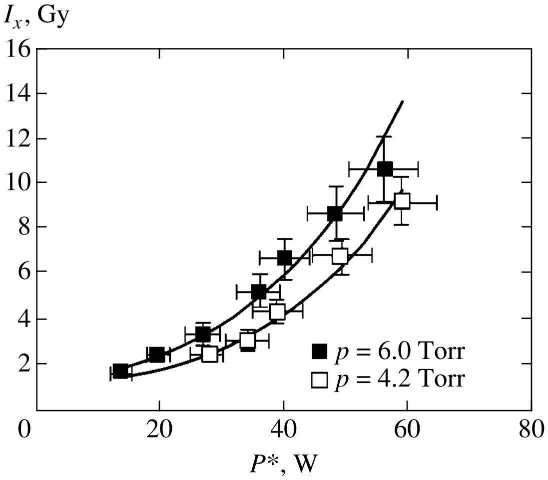


Fig. 14. Plots of the total dose of X-ray quanta emitted from a Ti cathode and detected for 6000 s by a TLD spaced by 7 cm from the cathode (with allowance for the detector efficiency) versus effective discharge power $P^* = UJQ$ ($Q = 0.15$ is the current pulse on-off ratio) for two gas pressures: $p_1 = 6.0$ Torr, and $p_2 = 4.2$ Torr.

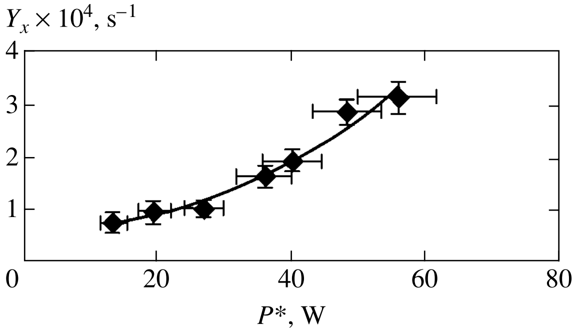


Fig. 15. Plot of the yield of X-ray quanta per deuteron versus effective discharge power P^* at a gas pressure of $p = 4.2$ Torr.

standard procedure of evaluation of the screening potential, including the normalization of the yields from bulk targets at various energies (with respect to the yield at the maximum energy) and comparison to predictions of the Bosch–Halle extrapolation, may exclude the errors related to uncertainties in the deuterium concentration and the absolute DD reaction yield.

The stopping power of deuterons in Ti in our experiments was assumed to be proportional to the particle velocity v_d. According to the results of recent investigations for metals and semiconductors [23, 24], this assumption is valid, at least for deuteron energies of $E_d \geq 1.0$ keV. The available data suggest that proportionality of the stopping power and velocity of deuterons is also not significantly distorted for $E_d \geq 0.5$ keV. Nevertheless, measurement of the stopping power for deuterons at $E_d \leq 1.0$ keV, which might reveal deviations from the $dE/dx \sim v_d$ relation, would be of considerable value.

In this study, the deuteron beam current was determined with an oscilloscope and, in fact, considered equal to the discharge current without using a Faraday cup. It is important to note, however, that this determination of the beam current cannot lead to overestimation of the DD reaction yield. Indeed, the secondary electron emission from the cathode, should it be caused by the bombardment with deuterons, might only reduce the total charge detected by the Faraday cup on the cathode (thus, understating the real deuteron beam current). In turn, a decrease in the deuteron current on the target may only lead to an increase in the absolute DD reaction yield. For this reason, estimates of the 3-MeV proton yield presented in Table 1 can be considered only as the most conservative values.

The phenomenon of deuteron channeling, which may increase the probability of deuteron collisions, was considered in detail in [6, 7]. Not dwelling here on this phenomenon in much detail, we only point out that channeling is not the main factor responsible for an increase in the DD reaction cross section because of the random orientation of polycrystals and the radiation damage of crystal structure in the Ti target, especially in the case of an intense flux of bombarding deuterons and a high target temperature. The effect of a molecular component (D_2^+/D_3^+) of the glow discharge on the DD reaction yield enhancement in Ti can be also ignored, to a first approximation, because the results of recent experiments with the beams of D^+ and D_3^+ showed analogous enhancement [6].

On the other hand, a large (as compared to the case of experiments using accelerators) uncertainty of the discharge conditions (voltage–current) and the temperature on the cathode surface can be actually considered as the main sources of errors in estimating the enhancement of the DD reaction yield in glow discharge. Indeed, the voltage–current fluctuations in our experiments might reach a level of 15% at a constant pressure of deuterium. These effects increase the experimental error in thick target yield and in evaluating the enhancement factor. As a result, the finite screening potential obtained using formula (3) is $U_e = 610 \pm 150$ eV, where the uncertainty ($\delta U_e = \pm 150$ eV) amounts to about 25% (including corrections for the current–voltage instability) of the total U_e value. At the same time, in the absence of significant spread in the deuteron energy in the experiments using accelerators, the error in calculations of the screening potential does not exceed 15%. It should be emphasized that, in such experiments at low deuteron energies ($E_d < 5$ keV), the uncertainty in the screening potential is determined primarily by the systematic error of detection, because a relatively small beam current leads to a rather low count rate for the DD reaction products. On the contrary, the count rate of 3-MeV protons in the case of glow discharge even at $E_d \leq 1$ keV is still greater than the statistical error of measurements, which is due to a high deuteron beam current on the cathode. In any case, the experimental DD yields in a Ti cathode are always much higher than the values obtained by extrapolation of the standard yield. As the deuteron energy decreases, the difference between experimental and standard yields tends to increase and remains significantly greater than all experimental uncertainties, including the detection errors and the corrections for the current/voltage instability.

In order to vary the deuteron energies, our experiments were performed with variable discharge voltages and currents (deuterium pressures) and, hence, with different discharge powers. For taking into account the effect of discharge power, which accounts for heating of the near-surface cathode layer (with a thickness comparable with the deuteron range in Ti) and the resulting change of the deuterium concentration, we introduced the correction coefficient $k(W, T)$ defined by formula (2) with normalization to the maximum power at the melting temperature of Ti. It should be noted that $k(W, T) = \varepsilon_d \Delta T / k_B T_m T_0 \approx 0.3$, and this value practically does not increase on changing the power. The validity of this approach is confirmed by the fact that the activation energy ε for deuteron escape from the Ti cathode surface (Fig. 5) was determined from the data determined in the experiments using accelerators at low temperatures (170–190 K) and can be considered as having the maximum value, since it has a tendency to decrease with increasing temperature [26]. Thus, a correction for possible changes in the deuterium concentration in Ti with allowance for the power supplied to the discharge is quite justified. In the absence of this correction, the experimental DD reaction yields at low deuteron energies would be unreasonably overstated. Therefore, the enhancement factor according to formula (3), with normalization to the effective cathode temperature according to formula (1), can be considered as a conservative estimate of the absolute DD reaction yield.

Table 2. Screening potentials and the corresponding electron energies for some metals

Target [Ref.]	$E_{d(lab)}$, keV	ΔI, mA	T, K	U_e, eV	Electron level	Level energy E, eV
Ti [7]	5–30	0.054	263	≤30	Ti (M_{II}/M_{III})	32.6
Ti [10]	2.5–10.0	0.06–0.25	186	65 ± 15	Ti (M_I)	58.3
Ti*	0.8–2.45	225–450	>1000	610 ± 150	Ti (L_{II})	461
Au [7]	5–30	0.054	263	61 ± 20	Au (O_{II})	71
Au [10]	2.5–10.0	0.06–0.25	180	70 ± 10	Au (O_{II})	71
Pd [7]	5–30	0.054	263	800 ± 70	Pd (M_I)	670
Pd [10]	2.5–10.0	0.06–0.30	313	310 ± 30	Pd (M_V)	334
PdO [10]	2.5–10.0	0.06–0.30	193	600 ± 20	Pd (M_{II})	560

* Experiment in glow discharge.

It should be noted that an additional factor influencing the DD reaction enhancement is related to the atomic vibrations in the target. The role of this factor must increase with temperature [24]. However, using published data [25], it can be readily shown that the corresponding enhancement correction for the glow discharge with $E_d = 1.0$ keV, and energy spread of $\Delta E \approx 100$ eV with the corresponding thermal ($E_{th} \sim 7.0 k_B T$) and vibrational ($E_{vib} \sim 0.1$ eV [22]) energies, does not exceed ±10%. This correction to the enhancement factor can increase the $f(E)$ value and must be considered as the limiting value. However, the real enhancement factor at $E_d = 1.0$ keV according to formula (3) is $f(E) \sim 10^9$. Therefore, neither a large energy spread nor a higher temperature in our experiments could significantly influence the DD reaction enhancement determined using formula (3).

In concluding the analysis of correctness of the DD reaction enhancement data obtained in our experiments with glow discharge, it should be noted that the contribution of D(d, p)T reaction yield due to the possible cracking of metal deuterides is lower by at least two orders of magnitude [27] than that obtained in our experiments (without correction for the diffusion processes) even at a minimum discharge voltage (Table 1). Thus, despite uncertainties related to the instability of conditions in a high-current glow discharge as compared to those in a deuteron accelerator, the data on a considerable enhancement of the DD reaction yield (and the corresponding screening potential) obtained in this study appear as significantly well substantiated from the standpoint of both an analysis of possible errors and a simple explanation of the DD reaction enhancement effect.

It was established that the anomalous enhancement of the DD reaction yield in a glow discharge in deuterium with a Ti cathode is accompanied by intense X-ray emission from the cathode surface. Such emission was not observed earlier in the experiments using accelerators at low energies, which is probably explained by small deuteron beam currents and significantly lower power injected into a metal target bombarded with an accelerator as compared to the case of glow discharge.

Without aspiring to give a detailed description of processes on a Ti cathode surface under the conditions of high-current deuteron beam action, we will now briefly outline a qualitative model capable of explaining the observed simultaneous DD reaction yield enhancement and the intense emission of soft X rays from the cathode. The mechanism of electron screening in metals at low energies of bombarding particles considered within the framework of the usual model of valence electron screening [28, 29] is still not completely clear and there is no commonly accepted explanation of the enhancement effect corresponding to a screening potential of $U_e > 100$ eV (i.e., well above the adiabatic limit). The presence of a high screening potential at high temperatures of the target with simultaneous intense X-ray emission cannot be explained in terms of classical Debye screening (in application to quasi-free valence electrons) or dynamic deuteron screening in the metal lattice [8, 30]. In this respect, we may only suggest that electrons of the inner levels can be also involved in screening [31].

In order to illustrate this hypothesis, we summarized in Table 2 the data on the screening potentials in some metal targets bombarded with deuterons. These data were mostly obtained in experiments using accelerators [6–10], but also include the results obtained for glow discharge in this study. The experimental values of the screening potential in Table 2 are compared to the energies of the closest inner electron shells of the corresponding metals. As can be seen from these data, the screening potentials reported by various researchers are in fact close (to within 1–2 standard deviations) to the energies of the inner electron shells of these target atoms.

The correlation between the electron screening potentials U_e for deuterons and the inner energy levels of the target atoms stipulated a relation between the mechanisms of this screening and the X-ray emission. One possible explanation for the strong electron screening of deuterons in Ti assumes that the bombarding deuterons coherently interact with a diffusing deuterium flux in the near-surface layer of the Ti cathode [26, 30, 32]. The high mobility of deuterium results from a high density of the deuteron beam current and the related high temperature of the target surface layer (with a thickness comparable with the deuteron range). Deuterons diffusing in *d*-metals possessing a high affinity to hydrogen must exhibit interactions with the valence *d* orbitals of metal atoms, that is, the *sd* hybridization of the valence electrons of deuterium and titanium on the outer electron shells [30].

For simplicity, we suggest that the interaction of an impinging deuteron with the valence *d* shell of Ti (coupled to the *s* orbital of another deuteron diffusing with an energy of $k_B T$, or practically resting relative to the projectile) has an inelastic character. The projectile energy is transmitted to a linear combination of wavefunctions corresponding to the given set of states of the diffusing deuteron coupled to the *d* orbital of Ti. This energy transfer leads to a nonzero probability of population of the inner (L_{II}) shell of titanium by the recoil deuteron and a temporal electron transition from the L_{II} shell to a state with lower energy [33]. As a result, the interaction between the resting deuteron (coupled to the valence *d* shell of Ti) and the recoil deuteron (occurring on the L_{II} shell with a binding energy of 462 eV for a very short time $\tau \approx \hbar/E_d \approx 10^{-18}$ s [34, 35]). This provides the conditions for strong screening, while the energy transferred from the deuteron to the L_{II}-shell electron is emitted in the form of an X-ray quantum [35]. The X-ray quantum energy might be approximately equal to that of the impinging deuteron, suggesting that the interaction would be fully inelastic. The process of simultaneous screening and X-ray emission involving the L_{II} shell of Ti can be described using the following scheme:

$$\begin{aligned} &D^+(E_d) + [D_0 + \mathrm{Ti}]_{sd} \\ &\longrightarrow [D_r^+ + \mathrm{Ti}(L_{II})]^* + D_0 \longrightarrow dd(U_{L_{II}}) + X(E_d), \end{aligned} \tag{5}$$

where $D^+(E_d)$ is the impinging deuteron with the energy E_d; $[D_0 + \mathrm{Ti}]_{sd}$ is the diffusing or resting deuteron with the energy $k_B T$ (k_B is the Boltzmann constant), which is coupled to the valence *d* shell of Ti (*sd* hybridization); $[D_r^+ + \mathrm{Ti}(L_{II})]^*$ is the bound excited state of the recoil deuteron D_r^+ and an L_{II} electron in Ti; $dd(U_{L_{II}})$ is the DD reaction between D_r^+ and D_0 with the screening potential $U_e = U_{L_{II}} = 462$ eV; and $X(E_d)$ is the emission of X-ray quantum with the energy E_d.

Under the glow discharge conditions, both the concentration and mobility of deuterium in a near-surface layer of Ti are much higher than the corresponding values in the experiments using accelerators, which is related to a high deuteron beam current density in the discharge. Therefore, the probability of finding a diffusing deuteron near a valence orbital of Ti in a glow discharge is also high. This circumstance gives rise to the probability of a quasi-inelastic collision between the diffusing and impinging deuterons, which allows the projectile to penetrate into the inner (L_{II}) shell of Ti. In the case of experiments using accelerators, the lower concentrations and mobilities of deuterons in the target lead to predominantly elastic collisions characterized by a lower energy transferred to a deuteron interacting with the valence orbital of Ti. As a result, the recoil deuterons (possessing much lower effective kinetic energies as compared to those of the impinging deuterons) cannot penetrate into the *L* shell of Ti. Should such deuterons reach the *M* shell, their screening will take pace at a lower electron screening potential ($U_e \sim 30$–65 eV), in agreement with experiment (see Table 2).

The approach outlined above for explaining the simultaneous screening of deuterons and X-ray quanta emission in glow discharge agrees with the observed dependence of the X-ray quantum yield Y_x on the effective discharge power P^* (Fig. 15). The yield of X-ray quanta with an energy of 1.25 ± 0.25 keV in the range of voltages 900–1700 V grows exponentially with the effective power. Since the effective temperature in a near-surface layer of Ti is proportional to the discharge power, the mobility of deuterium in this layer is also proportional to P^*. Thus, the simultaneous increase in the X-ray quanta emission intensity and the deuterium mobility in the near-surface layer of cathode (target) agrees with the model according to scheme (5).

5. CONCLUSIONS

We have studied for the first time the DD reaction yield (3-MeV protons) and the soft X-ray quanta emission from a titanium (Ti) cathode surface in a periodic pulsed glow discharge in deuterium at a discharge voltage of 0.8–2.45 kV and a discharge current density of 300–600 mA/cm^2. An analysis of the DD reaction yield as a function of the accelerating voltage allowed us to estimate the electron screening potential at $U_e = 610 \pm 150$ eV in the range of deuteron energies $0.8\ \mathrm{keV} < E_d < 2.45$ keV. The data obtained show evidence for a significant enhancement of the DD reaction yield in Ti in comparison to both theoretical estimates (based on the extrapolation of the known DD reaction cross section for $E_d \geq 5$ keV to low deuteron energies in the Bosch–Halle approximation) and the results of experiments using accelerators at deuteron energies of $E_{lab} \geq 2.5$ keV and current densities of 50–500 μA/cm^2. Intense emission of X-ray quanta (10^{13}–10^{14} s^{-1} cm^{-2}) was observed at an average energy of 1.2–1.5 keV.

Irrespective of a real mechanism responsible for the DD reaction yield enhancement in a glow discharge with a Ti cathode, accompanied by soft X-ray emission, we showed that a high-current pulsed glow discharge offers a unique possibility for studying the screening of low-energy deuterons by electrons and the accompanying atomic processes in solids under the conditions of a large energy density transfer. In order to increase the accuracy of data related to calculations of the DD reaction yields under glow discharge conditions, we plan to measure the energy distributions and charged states of deuterium using methods developed for low-temperature plasma diagnostics. In addition, we will also perform direct measurements of the deuterium concentration in the cathode using in situ measurements of its electric resistance in the course of deuteron bombardment.

ACKNOWLEDGMENTS

The authors are grateful to V.V. Kushin and A.I. Per’kov (Moscow Institute of Engineering Physics) for their help in experimental setups. Special thanks to J. Kasagi and G. Hubler for their interest in this study and valuable remarks, and to E.L. Feinberg and A.N. Lebedev (Institute of Physics, Russian Academy of Sciences) for fruitful discussions.

This study was partly financed by the government of the USA (AF ERC grant no. RP50053).

REFERENCES

1. U. Greife, F. Gorris, M. Junker, *et al.*, Z. Phys. A **465**, 150 (1995).
2. M. Junker, A. D’Alessandro, S. Zavatarelli, *et al.*, Phys. Rev. C **57**, 2700 (1998).
3. H. Yuki, T. Sato, J. Kasagi, *et al.*, J. Phys. G: Nucl. Part. Phys. **23**, 23 (1997).
4. M. Aliotta, F. Raiola, G. Gyurky, *et al.*, Nucl. Phys. A **690**, 790 (2001).
5. K. Czerski, A. Huke, A. Biller, *et al.*, Europhys. Lett. **54**, 449 (2001).
6. F. Raiola, P. Migliardi, G. Gyurky, *et al.*, Eur. Phys. J. A **13**, 377 (2002).
7. F. Raiola, P. Migliardi, L. Gang, *et al.*, Phys. Lett. B **547**, 193 (2002).
8. F. Raiola *et al.*, Europhys. J. A **19**, 283 (2004).
9. H. Yuki, J. Kasagi, A. G. Lipson, *et al.*, JETP Lett. **68**, 823 (1998).
10. J. Kasagi, H. Yuki, T. Baba, *et al.*, J. Phys. Soc. Jpn. **71**, 2881 (2002).
11. H. S. Bosch and G. M. Halle, Nucl. Fusion **32**, 611 (1992).
12. V. M. Grebenyuk, S. S. Parzhitski, P. M. Penkov, *et al.*, Laser Part. Beams **18**, 325 (2000).
13. V. M. Bystritskii, V. M. Bystritsky, S. A. Chaikovsky, *et al.*, Phys. At. Nucl. **64**, 855 (2001).
14. V. M. Bystritsky, V. V. Gerasimov, A. R. Krylov, *et al.*, Phys. At. Nucl. **66**, 1683 (2003).
15. K. Krieger, Science **304**, 1226 (2004).
16. D. Reiter, Trans. Fusion Technol. **33**, 249 (1998).
17. A. B. Karabut, Ya. A. Kucherov, and I. B. Savvatimova, Phys. Lett. A **170**, 265 (1992).
18. V. Violante, A. Torre, G. Silvaggi, and G. H. Miley, Fusion Technol. **39**, 266 (2001).
19. E. P. Velikhov, A. S. Kovalev, and A. T. Rakhimov, *Physical Phenomena in a Gas Discharge Plasma* (Nauka, Moscow, 1987) [in Russian].
20. F. G. Baksht and V. G. Yuriev, Zh. Tekh. Fiz. **49**, 905 (1979) [Sov. Phys. Tech. Phys. **24**, 535 (1979)].
21. A. G. Lipson, A. B. Karabut, and A. S. Roussetski, Proc. Ital. Phys. Soc. **70**, 335 (2001).
22. H. H. Anderson and J. F. Ziegler, *Hydrogen Stopping Powers and Ranges in All Elements* (Pergamon, New York, 1977).
23. K. Eder, D. Semard, P. Bauer, *et al.*, Phys. Rev. Lett. **79**, 4112 (1997).
24. S. P. Moller, A. Csete, T. Ichioka, *et al.*, Phys. Rev. Lett. **88**, 193201 (2002).
25. L. Schlapbach, I. Anderson, and J. P. Burger, in *Materials Science and Technology*, Ed. by K. H. Jurgen Buschow (Weinheim, New York, 1994), Vol. 3B, Part 2.
26. Y. Fukai and H. Sugimoto, Adv. Phys. **34**, 263 (1985).
27. A. G. Lipson, B. F. Lyakhov, N. Asami, *et al.*, Fusion Technol. **38**, 257 (2000).
28. C. Bonomo, G. Fiorentini, Z. Fulop, *et al.*, Nucl. Phys. A **719**, 37C (2003).
29. G. Fiorentini, C. Rolf, F. L. Villante, and B. Ricci, Phys. Rev. C **67**, 014603 (2003).
30. S. Ichimaru, Rev. Mod. Phys. **65**, 255 (1993).
31. K. P. Shina and P. L. Hagelstein, Proc. Ital. Phys. Soc. **70**, 369 (2001).
32. D. Pines, *Elementary Excitations in Solids* (Wiley, New York, 1963; Mir, Moscow, 1965).
33. P. B. Corkum, Phys. Rev. Lett. **71**, 1994 (1993).
34. M. Drescher, M. Hentschei, R. Klenberger, *et al.*, Nature **419**, 803 (2002).
35. J. Seres, E. Seres, A. J. Verhoff, *et al.*, Nature **433**, 596 (2005).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 100, No. 6, 2005, pp. 1188–1233.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 127, No. 6, 2005, pp. 1350–1402.
Original Russian Text Copyright © 2005 by Suslov.

MISCELLANEOUS

Divergent Perturbation Series

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

Received January 25, 2005

Abstract—Various perturbation series are factorially divergent. The behavior of their high-order terms can be determined by Lipatov's method, which involves the use of instanton configurations of appropriate functional integrals. When the Lipatov asymptotic form is known and several lowest order terms of the perturbation series are found by direct calculation of diagrams, one can gain insight into the behavior of the remaining terms of the series, which can be resummed to solve various strong-coupling problems in a certain approximation. This approach is demonstrated by determining the Gell-Mann–Low functions in φ^4 theory, QED, and QCD with arbitrary coupling constants. An overview of the mathematical theory of divergent series is presented, and interpretation of perturbation series is discussed. Explicit derivations of the Lipatov asymptotic form are presented for some basic problems in theoretical physics. A solution is proposed to the problem of renormalon contributions, which hampered progress in this field in the late 1970s. Practical perturbation-series summation schemes are described both for a coupling constant of order unity and in the strong-coupling limit. An interpretation of the Borel integral is given for "non-Borel-summable" series. Higher order corrections to the Lipatov asymptotic form are discussed. © *2005 Pleiades Publishing, Inc.*

CONTENTS

1063-7761/05/10006-1188$26.00 © 2005 Pleiades Publishing, Inc.

1. DYSON'S ARGUMENT: IMPORTANT PERTURBATIVE SERIES HAVE ZERO RADIUS OF CONVERGENCE

Classical books on diagrammatic techniques [2–4] describe the construction of diagram series as if they were well defined. However, almost all important perturbation series are hopelessly divergent since they have zero radii of convergence. The first argument to this effect was given in [5] with regard to quantum electrodynamics. Here, it is reiterated by using simpler examples.

Consider a Fermi gas with a delta-function interaction $g\delta(\mathbf{r}-\mathbf{r}')$ and the corresponding perturbation series in terms of the coupling constant g. Its radius of convergence is determined by the distance from the origin to the nearest singular point in the complex plane and can be found as follows. In the case of a repulsive interaction ($g > 0$), the ground state of the system is a Fermi liquid. When the interaction is attractive ($g < 0$), the Cooper instability leads to superconductivity (see Fig. 1a). As g is varied, the ground state qualitatively changes at $g = 0$. Thus, the nearest singular point is located at the origin, and the convergence radius of the series is zero.

An even simpler example is the energy spectrum of a quantum particle in the one-dimensional anharmonic potential

$$U(x) = x^2 + gx^4. \tag{1.1}$$

Whereas the system has well-defined energy levels when $g > 0$, these levels are metastable when $g < 0$ since the particle can escape to infinity (see Fig. 1b). Therefore, the perturbation series in terms of g is divergent for any finite g. This can easily be shown by calculating its coefficients. The calculation of the first 150 coefficients in [6] served as direct evidence of convergence of the series and basis for a detailed convergence study.

Zero radius of convergence looks "accidental" in quantum-mechanical problems: it is the case when a potential of special form and an "inadequate" coupling constant are chosen deliberately. However, zero radius of convergence is encountered in all fundamental quantum field theories with a single coupling constant.

Even though Dyson's argument is unquestionable, it was hushed up or decried for many years, because the scientific community was not prepared to face the problem of divergent perturbation series.

2. LIPATOV'S METHOD: QUANTITATIVE ESTIMATION OF DIVERGENCE OF SERIES EXPANSIONS

A further step was made in 1977, when Lipatov's method was proposed as a tool for calculating high-order terms in perturbation series and making quantitative estimates for the divergence of series. The idea of the method is as follows. If a function $F(g)$ can be

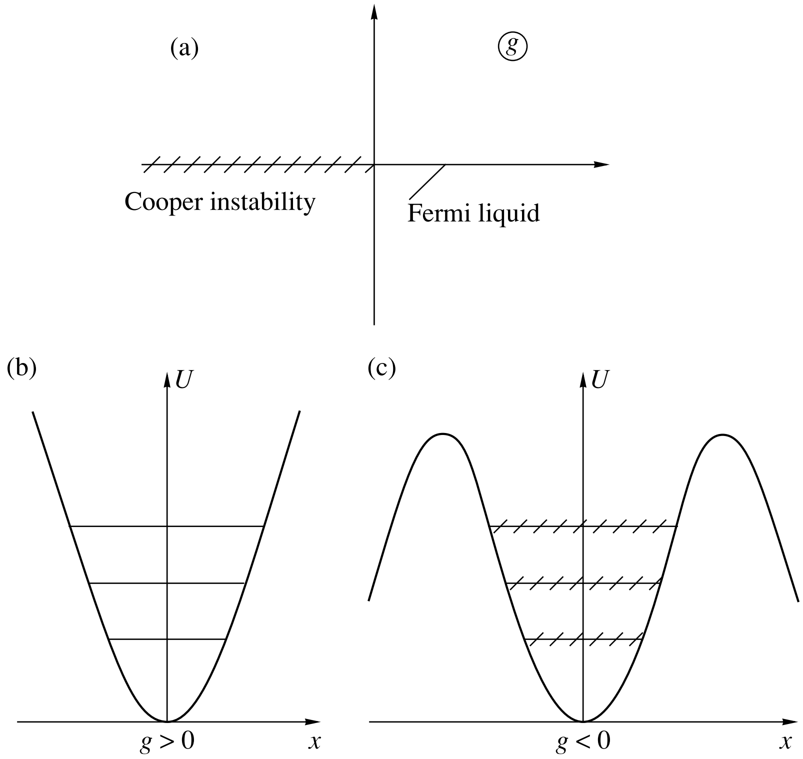


Fig. 1. Graphic illustration of Dyson's argument.

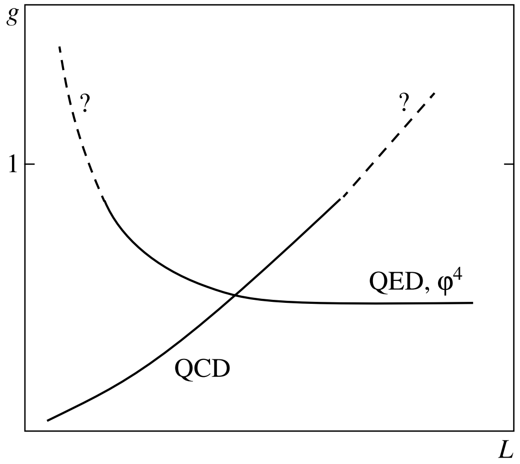


Fig. 2. Effective coupling versus length scale in φ^4 theory, QED, and QCD.

expanded into a series in terms of g,

$$F(g) = \sum_{N=0}^{\infty} F_N g^N, \tag{2.1}$$

then the coefficients F_N in the expansion can be determined as

$$F_N = \int_C \frac{dg}{2\pi i} \frac{F(g)}{g^{N+1}}, \tag{2.2}$$

where the contour C goes around the point $g = 0$ in the complex plane. Rewriting the denominator as $\exp\{-(N+1)\ln g\}$ for large N, one may hope that the saddle-point method can be applied to the resulting exponential with a large exponent.

It is well known that the problems tractable by diagrammatic techniques can be reformulated in terms of functional integrals of the form

$$Z(g) = \int D\varphi \exp(-S_0\{\varphi\} - gS_{\text{int}}\{\varphi\}), \tag{2.3}$$

with the expansion coefficients

$$\begin{aligned} Z_N &= \int_C \frac{dg}{2\pi i g} \\ &\times \int D\varphi \exp(-S_0\{\varphi\} - gS_{\text{int}}\{\varphi\} - N\ln g). \end{aligned} \tag{2.4}$$

Lipatov suggested seeking the saddle-point configuration in (2.4) in g and φ simultaneously, rather than with respect to g only. The desired configuration exists in all cases of interest, being defined on a localized function $\varphi(x)$ called *instanton.* However, it turns out that the saddle-point approximation is applicable when N is large irrespective of its applicability to integral (2.3). This finding has important consequences: whereas functional integrals cannot generally be calculated exactly, they can always be calculated in the saddle-point approximation.

Once expansion coefficients are known for a functional integral, expansions of Green functions, vertices, etc., can be found, because factorial series can be treated as finite expressions in a simple algebra (see [8, Section 5.3]). Generally, the Lipatov asymptotic form of the expansion coefficients for any quantity $F(g)$ is

$$F_N = c\Gamma(N+b)a^N, \tag{2.5}$$

where $\Gamma(x)$ is the gamma function and a, b, and c are parameters depending on the specific problem under analysis. In the framework of a particular theory, a is a universal constant, b is a parameter depending on $F(g)$, and c depends on external coordinates or momenta.

When the Lipatov asymptotic form is known and a few lowest order terms of a perturbation series are found by direct calculation of diagrams, one can gain insight into the behavior of the remaining terms of the series and perform their summation to solve various strong-coupling problems in a certain approximation. The most important consequence is the possibility of finding the Gell-Mann–Low function $\beta(g)$, which determines the effective coupling constant $g(L)$ as a function of length scale:

$$-\frac{dg}{d\ln L^2} = \beta(g). \tag{2.6}$$

In relativistic theories, the first term in the expansion of $\beta(g)$ is quadratic, $\beta(g) = \beta_2 g^2 + \ldots$. For a small g, Eq. (2.6) yields the well-known result [3, 9, 10]

$$g(L) = \frac{g_0}{1 - \beta_2 g_0 \ln(L^2/L_0^2)}, \tag{2.7}$$

where g_0 is the value of $g(L)$ on a length scale L_0. In both quantum electrodynamics (QED) and φ^4 theory, the constant β_2 is positive, and $g(L)$ is an increasing function at small L (see Fig. 2). In quantum chromodynamics (QCD), the sign of β_2 is negative. Accordingly, the interaction between quarks and gluons is weak at small L (asymptotic freedom), while its increase with L demonstrates a tendency toward confinement (see Fig. 2). One problem of primary importance is extension of (2.7) to intermediate and strong coupling. According to the classification put forward in [2], the function $g(L)$ tends to a constant if $\beta(g)$ has a zero at a finite g and continues to increase ad infinitum as $g \longrightarrow \infty$ if $\beta(g) \propto g^\alpha$ with $\alpha \le 1$. If $\beta(g) \propto g^\alpha$ with $\alpha > 1$, then two interpretations are plausible. On the one hand, assuming finite interaction at long distances, one would have a self-contradictory theory: the effective charge

$g(L)$ goes to infinity at a finite L_c (Landau pole), while the function $g(L)$ is undefined at $L < L_c$. On the other hand, a field theory interpreted as a continuum limit of lattice models is "trivial" since the interaction vanishes as $L \longrightarrow \infty$ ("zero-charge" property). The first attempts to determine the Gell-Mann–Low function in φ^4 theory were made in [11–13].

Originally developed for scalar theories (such as φ^4 [7]), Lipatov's method was extended to vector fields [14], fermion problems [15], scalar electrodynamics [16], and the Yang–Mills theories [17, 18], as well as to a variety of problems in quantum mechanics (see [19] and reviews in [20, 21]). Next in order were its applications to theories of practical interest, QED [22, 23] and QCD [24–26].

In all theories mentioned above, factorially divergent series were obtained. Assuming that divergent series are "the devil's invention," one must admit that the Creator has also taken part: modeling of physical reality leads to divergent series expansions with striking regularity.

3. INTERPRETATION OF PERTURBATION SERIES: A SURVEY OF THE MATHEMATICAL THEORY OF DIVERGENT SERIES

The modern status of divergent series suggests that techniques for manipulating them should be included in a minimum syllabus for graduate students in theoretical physics. However, the theory of divergent series is almost unknown to physicists, because the corresponding parts of standard university courses in calculus date back to the mid-nineteenth century, when divergent series were virtually banished from mathematics. The discussion that follows provides a brief review of the mathematical theory of divergent series [27].

3.1. Can We Deal with Divergent Series?

Dealing with series of the form

$$a_0 + a_1 + a_2 + a_3 + \ldots + a_N + \ldots \tag{3.1}$$

for the first time, one may be tempted to treat them as if they were finite sums. However, this is incorrect in the general case, because a series can be treated as a finite sum only if it is absolutely convergent [28], i.e.,

$$|a_0| + |a_1| + |a_2| + |a_3| + \ldots + |a_N| + \ldots < \infty. \tag{3.2}$$

When dealing with a conditionally convergent series, such as the alternating harmonic series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \ldots, \tag{3.3}$$

one cannot rearrange its terms in an arbitrary manner: by Riemann's theorem, a conditionally convergent series can be rearranged to converge to any specified sum [28]. Indeed, the sum of a convergent series is defined as the limit of its partial sums, and any result can be obtained by shifting negative terms rightwards and positive terms leftwards, or vice versa.

Expectably, the analysis of divergent series is even more complicated because of a greater number of forbidden operations on them:

(a) obviously, terms cannot be rearranged;

(b) terms cannot be grouped either, e.g.,

$$\begin{aligned}1 - 1 + 1 - 1 + 1 - 1 + \ldots &\neq (1-1) + (1-1)\\ + (1-1) + \ldots &\neq 1 + (-1+1) + (-1+1)\\ &+ (-1+1) + \ldots;\end{aligned} \tag{3.4}$$

(c) a series cannot be "padded" by inserting zero terms,

$$\begin{aligned}&a_0 + a_1 + a_2 + a_3 + \ldots\\ &\neq a_0 + 0 + a_1 + 0 + a_2 + 0 + a_3 + 0 + \ldots.\end{aligned} \tag{3.5}$$

Thus, the basic idea of the theory of divergent series can be stated: in principle, they can be consistently manipulated if one follows rules that are much more stringent than those for operations on finite sums or convergent series.

3.2. Euler's Principle

What are the new rules to be followed? A preliminary answer to this question was given by L. Euler, who was the true pioneer in developing the theory of divergent series. Euler ruled out the use of number series (3.1) and expansions over arbitrary basis functions,[1]

$$a_0 f_0(x) + a_1 f_1(x) + a_2 f_2(x) + \ldots + a_N f_N(x) + \ldots, \tag{3.6}$$

and emphasized a special role played by power series

$$a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots + a_N x^N + \ldots. \tag{3.7}$$

Power series expansions are special in that information about the natural numbering of their terms is preserved under permutation or other operations. Therefore, power series can be treated as finite sums. It is clear that forbidden operations are ruled out automatically: if number series (3.1) is interpreted as the limit of power series (3.7) as $x \longrightarrow 1$, then any rearrangement, pad-

---

[1] This discussion concerns to divergent series only. Convergent expansions such as (3.6) (e.g., over an orthogonal basis) are obviously admissible.

ding with zero terms, or association leads to a series different from the starting one:

$$\begin{gathered}a_1 + a_0 + a_3 + a_2 + \ldots \\ \longrightarrow a_1 + a_0 x + a_3 x^2 + a_2 x^3 \ldots,\end{gathered} \tag{3.8}$$

$$\begin{gathered}a_0 + 0 + a_1 + 0 + a_2 + 0 + a_3 + \ldots \\ \longrightarrow a_0 + 0 \cdot x + a_1 x^2 + 0 \cdot x^3 + a_2 x^4 \\ + 0 \cdot x^5 + a_3 x^6 \ldots,\end{gathered} \tag{3.9}$$

$$\begin{gathered}(a_0 + a_1) + (a_2 + a_3) + (a_4 + a_5) + \ldots \\ \longrightarrow (a_0 + a_1) + (a_2 + a_3)x + (a_4 + a_5)x^2 + \ldots.\end{gathered} \tag{3.10}$$

The fundamental reason for Euler's principle lies in the fact that a power series is absolutely convergent within its circle of convergence and defines an analytic function that can be continued outside its domain of convergence. Accordingly, arbitrary manipulations of power series are admissible either as operations on absolutely convergent series or by the principle of analytic continuation. However, an analytic function may have several branches, and information about them is lost when divergent series are employed. Therefore, Euler's approach is not complete, and its application may lead to poorly defined expressions requiring correct interpretation. As a consequence, its rigorous mathematical substantiation is hampered by difficult problems. Generally, constructive results concerning divergent series provide partial evidence in support of Euler's principle under some restrictive assumptions [27]. Typically, the approach is valid for the entire parameter space spanned by the coefficients of a series except for a set of measure zero, where it is valid only when the definition of the sum of the series is appropriately generalized. The excluded set has a complex structure and is difficult to specify by proving only a finite number of theorems. For this reason, Euler's principle cannot be adopted in modern mathematics without reservation. However, it is not rejected either, because it does not seem to be disproved by any known fact.

Basically, Euler's principle is consistent with common practice in theoretical physics. It is commonly believed that formal manipulations of power series on a "symbolic" level cannot lead to results that are outright incorrect even if divergent series are used in intermediate calculations. Moreover, ill-defined expressions do not present significant problems, since their correct interpretation can be found from physical considerations by applying various rules for avoiding singularities, which are so skillfully devised by physicists. When applying this approach, one should follow two rules: never substitute the numerical values of x before the series is transformed into a convergent one and never perform Taylor series expansion about any known singular point.

With regard to the latter requirement, note that the series used in quantum field theories have zero radii of convergence, but arise from functional integral (2.3) as a result of a regular expansion of the exponential in terms of g and a subsequent (incorrect) interchange of summation and integration. In essence, a summation procedure is a reverse rearrangement performed "elsewhere": its applicability guarantees the freedom of formal manipulation.

It may seem that the restriction to power series expansion is very stringent. Actually, this is not true, because a number series may arise in a physical application only when some particular values are assigned to model parameters. Normally, a power series in at least one parameter can be obtained by returning to the general statement of the problem or generalizing the model. This is frequently done by using relatively simple tricks. For example, if the potential energy in the Schrödinger equation is treated as a perturbation, then the resulting expansion is not a power series. However, if $U(x)$ is replaced with $gU(x)$ before performing the series expansion (with a view to setting $g \longrightarrow 1$ as a final step), then a power series in g will be obtained.

Now, a few words should be said about number series (3.1). In principle, it can be consistently manipulated [27] if (3.1) is interpreted as a symbolic representation that cannot be treated as a conventional sum (otherwise, one is led to paradoxes commonly discussed in textbooks [28]). Manipulations of this kind are performed according to ad hoc rules known only to specialists. The constructive prescriptions in these rules are consistent with Euler's principle if number series (3.1) is identified with power series (3.7) in the limit of $x \longrightarrow 1$. This can always be done formally, but one must be sure that series (3.1) has not been modified by rearranging terms, discarding zero terms, etc. Since the fulfillment of this requirement cannot be reliably checked unless the number series is derived from a known power series, number series per se are of no practical importance.

As an implementation of Euler's approach, consider the well-known Borel transform: dividing and multiplying each term of a series by $N!$, introducing the integral representation of the gamma function, and permuting sums and integrals, one obtains

$$\begin{aligned}F(g) &= \sum_{N=0}^{\infty} F_N g^N = \sum_{N=0}^{\infty} \frac{F_N}{N!} \int_0^{\infty} dx x^N e^{-x} g^N \\ &= \int_0^{\infty} dx e^{-x} \sum_{N=0}^{\infty} \frac{F_N}{N!} (gx)^N.\end{aligned} \tag{3.11}$$

The power series on the right-hand side has factorially improved convergence properties and defines the Borel transform $B(z)$ of $F(g)$:

$$F(g) = \int_0^{\infty} dx e^{-x} B(gx), \quad B(z) = \sum_{N=0}^{\infty} \frac{F_N}{N!} z^N. \tag{3.12}$$

The Borel transform provides a natural basis for summing factorially divergent series in quantum field theories.

3.3. How Should We Define the Sum of a Series?

Now, let us discuss the modern approach to the problem. An "ideal" program of formalization can be represented as follows.

1. Formulate a definition of the sum S of a series that is equivalent to the conventional definition in the case of a convergent series.

2. Consider a class L of transformations of one series into another that leaves the sum invariant:

$$\begin{gathered} a_0 + a_1 + a_2 + a_3 + \ldots + a_N + \ldots = S \\ \longrightarrow b_0 + b_1 + b_2 + b_3 + \ldots + b_N + \ldots = S. \end{gathered} \tag{3.13}$$

3. Verify that L is a sufficiently wide class that can be used to transform convergent series into divergent ones and vice versa.

4. Specify the class L to set rules for manipulating series expansions without checking their convergence.

Has this program ever been implemented in modern mathematics? In fact, it has been, to the extent that a subclass of L sufficiently wide to solve practical problems has been specified. However, the theory cannot be presented in the elegant form outlined above. Indeed, difficulties arise even in implementing the first step: no definition of a sum equally suited to any particular problem is available. Summation methods for strongly divergent series are not instrumental as applied to weakly divergent ones, and vice versa. For this reason, a laissez-faire approach is adopted: any definition of a sum is formally admissible, and mutual consistency of different definitions is the only subject to be analyzed on an abstract mathematical level. The choice of a particular definition is left to the user. This attitude of mathematicians is not quite correct: the user that knows the definition of the sum can carry out the rest of the analysis. However, this attitude is well grounded (see Section 7).

In principle, it is known how these difficulties should be resolved. Recall the standard definitions of temperature in physics. Since no temperature measurement method is universally applicable, several temperature standards have been introduced (for high, low, and ultralow temperatures) that lead to identical results in the temperature regions where they overlap. An analogous approach can be adopted in the theory of divergent series, where a variety of "good" (mutually consistent) summation methods are available:[2] a sum can be defined by using a combination of these methods. Since good methods actually rely on Euler's principle, this approach reverts to that principle, but on a higher formal level and under certain restrictions.

As examples, consider the following possible definitions of sum.

Euler's definition. If power series (3.7) is convergent at small x, then it defines a regular function $f(x)$ whose analytic continuation is the sum of series (3.7) outside its circle of convergence.

In physical applications, this definition is adopted without reservation. As noted above, it is not complete, because the choice of a branch of the analytic function remains an open question. However, when this definition is meaningful, all calculations can be performed only by using convergent series, and the uncertainty is thus eliminated. A theory of divergent series is really necessary when the radius of convergence is zero, i.e., when Euler's definition is meaningless.

Borel's definition (applicable in the latter case as well). The sum of series (3.11) is given by (3.12). This definition agrees with other definitions based on Euler's principle and satisfies all necessary requirements.

3.4. Asymptotic Interpretation of Divergent Series

Modern theory of divergent series has "two sources and two parts." The foregoing discussion deals with the essentials of the Borel summation theory presented in its complete form in [29]. More widely known is the asymptotic interpretation of divergent series proposed by Poincaré [30]. A power series expansion of a function $f(x)$ is asymptotic if

$$f(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_N x^N + R_N(x), \tag{3.14}$$

where

$$R_N(x) = O(x^{N+1}), \quad x \longrightarrow 0, \tag{3.15}$$

i.e., if $f(x)$ is accurately approximated by a truncated series when x is small. The asymptotic interpretation is constructive only if the problem at hand involves a small parameter. However, when this condition is satisfied, one need not sum any high-order terms and even should not be interested in their behavior. Another advantage lies in the possibility of constructing asymptotic expansion (3.6) over arbitrary basis functions, provided that each $f_n(x)$ approaches zero faster than does $f_{n-1}(x)$.[3]

There is no one-to-one correspondence between functions and asymptotic power series expansions, because $f(x)$ can be modified by adding a function for which all coefficients in (3.14) vanish, such as

---

[2] Note that "poor" (mutually inconsistent) methods are no less abundant.

[3] If $x \geq 1$, then expansion (3.6), in contrast to (3.7), does not admit any meaningful interpretation at all. However, this is not necessary: why should we use an expansion that is not regular and does not involve any small parameter?

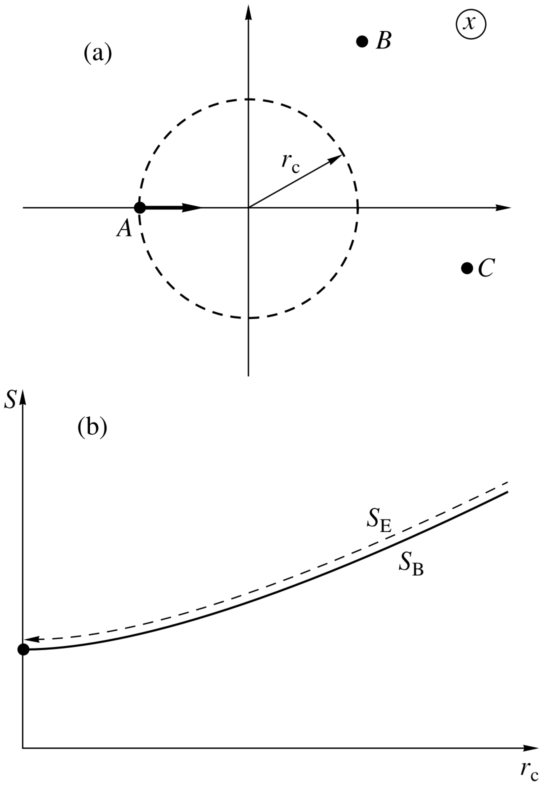


Fig. 3. (a) Convergence radius r_c decreases as singular point A approaches the origin. (b) Euler sum S_E equals Borel sum S_B when r_c is finite; when $r_c = 0$, the former is meaningless, whereas the latter extends the former by continuity.

$\exp(-a/x)$. When a series is convergent at small x, this uncertainty is eliminated by imposing the condition of analyticity at $x = 0$. However, this cannot be done for series with zero radii of convergence. At first glance, this would imply that a divergent series cannot be assigned any particular sum.

Actually, this is not true. Asymptotic equality means that the partial sum $a_0 + \ldots + a_N x^N$ "resembles" the function $f(x)$ up to some remainder term $R_N(x)$. It is no surprise that many functions meet this requirement within the prescribed accuracy. Their variety can be reduced by using a smaller $R_N(x)$, so that ultimately a single function remains. In fact, this is exactly what should be done: when the remainder term is subject to certain constraints, the function and the asymptotic series expansion are in one-to-one correspondence. Namely, it suffices to replace standard condition (3.15) with the so-called strong asymptotic condition

$$|R_N(x)| < C_N|x|^{N+1}, \quad x \in G, \tag{3.16}$$

where G is a region containing the point $x = 0$ and C_N is a specially chosen coefficient. It is reasonable to treat the single function that satisfies the strong asymptotic condition and "resembles" an asymptotic series "most closely" as its sum. By Watson's theorem on Borel summability [27], this unique function is given by the Borel integral in (3.12) for a broad class of divergent series. Thus, both summation theory and asymptotic theory naturally lead one to adopt Borel's sum as the sum of a divergent series.

The discussion above clearly solves the problem of nonperturbative contributions, such as $\exp(-a/x)$, which is frequently brought up as an argument against the use of perturbation series. When Borel's definition is adopted, addition of such terms to Borel integral (3.12) is forbidden. Formal manipulations of power series expansions will never lead to terms of this kind unless Taylor series expansion is performed about a known singular point.

3.5. Physical Arguments

Now, let us discuss physical arguments in support of the interpretation of perturbation series in the Borel sense.

Suppose that power series (3.7) has a finite radius of convergence. Then, the corresponding analytic function $f(x)$ has singular points A, B, C, … at finite distances from the origin (see Fig. 3a). In this case, it can be shown that Borel's definition of a sum is equivalent to Euler's, which is definitely suitable for physical applications.

As the distance from a pole or power-like singularity at a point A to the origin decreases, the coefficients in the series diverge, and the expansion becomes meaningless when the convergence radius approaches zero. However, there exist singularities that can be moved to the origin without causing divergence of the coefficients: these are branch points with exponentially decreasing jump across a cut. In particular, Lipatov asymptotic form (2.5) is associated with the following jump in $F(g)$ across a cut [31, 32]:[4]

$$\Delta F(g) = 2\pi i c\left(\frac{1}{ag}\right)^b \exp\left(-\frac{1}{ag}\right). \tag{3.17}$$

Both Borel's and Euler's sums vary with radius of convergence r_c, remaining equal. When $r_c = 0$, Euler's sum is meaningless, whereas Borel's sum corresponds to Euler's definition extended by continuity (see Fig. 3b).

The limit of $r_c \longrightarrow 0$ is amenable to a straightforward physical interpretation. Consider the Fermi gas with a delta-function interaction discussed in Section 1. The Cooper instability occurs for arbitrary $g < 0$ only at zero temperature. As the temperature T is raised to a finite value, the instability domain shifts to negative g by an amount g_c determined by a Bardeen–Cooper–

---

[4] Correspondence between (3.17) and (2.5) can be established by calculating the jump across the cut for Borel's sum of a series with expansion coefficients having asymptotic form (2.5). Alternatively, one can write Cauchy's integral formula for a point g lying in the domain of analyticity and deform the integration contour so that it goes around the cut. Then, the jump across the cut given by (3.17) will correspond to asymptotic expression (2.5).

Schriffer-like relation, $T \propto \exp\{-\mathrm{const}/g_c\}$. The corresponding perturbation series has a finite convergence radius g_c, which tends to zero with decreasing temperature. Generally, the value of a quantity calculated at strictly zero temperature differs from its limit value approached as $T \longrightarrow 0$. However, the physically meaningful value is that obtained as $T \longrightarrow 0$. Thus, the value at $T = 0$ must always be defined by continuous extension, i.e., by Borel summation.

4. LIPATOV ASYMPTOTIC FORMS FOR SPECIFIC MODELS

The calculation of Lipatov asymptotic form (2.5) is tedious if all parameters a, b, and c are to be found. However, its functional form can be found by performing a formal saddle-point expansion and separating the dependence on N. In what follows, calculations of this kind are performed for several fundamental models in theoretical physics.

4.1. φ^4 Theory

To begin with, consider the n-component φ^4 theory. The corresponding action is

$$S\{g, \varphi\} = \int d^d x \left\{ \frac{1}{2}\sum_{\alpha=1}^{n} [\partial_\mu \varphi_\alpha(x)]^2 + \frac{1}{2} m^2 \sum_{\alpha=1}^{n} \varphi_\alpha^2(x) + \frac{1}{4} g \left(\sum_{\alpha=1}^{n} \varphi_\alpha^2(x) \right)^2 \right\} \tag{4.1.1}$$

(d is the space dimension). Functional integrals of the form

$$Z_M(g) = \int D\varphi \varphi_{\alpha_1}(x_1)\varphi_{\alpha_2}(x_2)\dots\varphi_{\alpha_M}(x_M) \times \exp(-S\{g, \varphi\}) \tag{4.1.2}$$

define M-point Green functions,

$$G_M(g) = \frac{Z_M(g)}{Z_0(g)}, \tag{4.1.3}$$

which are diagrammatically represented by M-legged graphs. Hereinafter, integral (4.1.2) is written in compact form as

$$Z(g) = \int D\varphi \varphi^{(1)}\dots\varphi^{(M)} \exp(-S\{g, \varphi\}), \tag{4.1.4}$$

and normalized to an analogous integral with $M = 0$ and $g = 0$, with the factor $Z_0^{-1}(0)$ subsumed under $D\varphi$. Actually, the explicit form of the action is not required in the present analysis, and only its homogeneity is used to write

$$S\{g, \varphi\} = \frac{S\{\varphi\}}{g}, \quad \text{where} \quad \varphi = \frac{\phi}{\sqrt{g}}. \tag{4.1.5}$$

First, consider a finite-dimensional integral having the form of (4.1.4) with $D\varphi = d\varphi_1 d\varphi_2 \dots d\varphi_m$ and define

$$\phi = \begin{vmatrix} \phi_1 \\ \phi_2 \\ \cdot \\ \phi_m \end{vmatrix}, \quad S'\{\phi\} = \begin{vmatrix} \partial S/\partial\phi_1 \\ \partial S/\partial\phi_2 \\ \dots \\ \partial S/\partial\phi_m \end{vmatrix}, \quad S''\{\phi\} = \left\| \frac{\partial^2 S}{\partial\phi_i \partial\phi_j} \right\|. \tag{4.1.6}$$

This notation makes it possible to write any expression in a form analogous to the corresponding one-dimensional integral. In the infinite-dimensional limit, $S'\{\phi\}$ and $S''\{\phi\}$ become the first and second functional derivatives, interpreted as a vector and linear operator, respectively.

According to Section 2, the expansion coefficients are

$$Z_N = \oint_C \frac{dg}{2\pi i g} \int D\varphi \varphi^{(1)}\dots\varphi^{(M)} \times \exp\left(-\frac{S\{\phi\}}{g} - N \ln g \right), \tag{4.1.7}$$

and the saddle-point conditions have the form

$$S'\{\phi_c\} = 0, \quad g_c = \frac{S\{\phi_c\}}{N}. \tag{4.1.8}$$

The expansion of the exponent in (4.1.7) to quadratic terms in $\delta\phi = \phi - \phi_c$ and $\delta g = g - g_c$ is

$$-N - N\ln g_c - \frac{N}{2}\frac{(\delta\phi, S''\{\phi_c\}\delta\phi)}{S\{\phi_c\}} - \frac{N}{2g_c^2}(\delta g).^2 \tag{4.1.9}$$

Since

$$\delta\phi = \sqrt{g_c}\left(\delta\varphi + \frac{\delta g}{2g_c}\varphi_c \right), \quad \delta\varphi = \varphi - \varphi_c, \tag{4.1.10}$$

the origin of $\delta\varphi$ can be shifted to obtain

$$Z_N = e^{-N} g_c^{-N-M/2} \int_{-\infty}^{\infty} \frac{dt}{2\pi} \int D\varphi \phi_c^{(1)}\dots\phi_c^{(M)} \times \exp\left(-\frac{1}{2}(\delta\varphi, S''\{\phi_c\}\delta\varphi) + \frac{N}{2}t^2 \right), \tag{4.1.11}$$

where $\delta g = i g_c t$, because the saddle point is passed in the vertical direction. The Gaussian integration yields

$$Z_N = \frac{\mathrm{const}}{\sqrt{-\det[S''\{\phi_c\}]}} S\{\phi_c\}^{-N} \Gamma\left(N + \frac{M}{2} \right); \tag{4.1.12}$$

i.e., Lipatov asymptotic form (2.5) is recovered.

The result given by (4.1.12) is independent of m, remaining valid as $m \longrightarrow \infty$, i.e., in the functional-integral limit. However, any realistic functional integral contains zero modes associated with the symmetry of action under a continuous group defined by the operator $\hat{L}$, $S\{\phi\} = S\{\hat{L}\phi\}$. If ϕ_c is an instanton (i.e., $S'\{\phi_c\} = 0$), then so is $\hat{L}\phi_c$ ($S'\{\hat{L}\phi_c\} = 0$). By the continuity of the group operation, there exists an operator $\hat{L}$ that is arbitrarily close to the identity operator, $\hat{L}_\epsilon = 1 + \epsilon\hat{T}$. Hence, it is clear that $\hat{T}\phi_c$ is the eigenvector of the operator $S''\{\phi_c\}$ associated with its zero eigenvalue. Therefore, $\det[S''\{\phi_c\}] = 0$, and expression (4.1.12) is divergent. However, its divergence is spurious, being due to the inapplicability of the Gaussian approximation to integrals over zero modes.

To calculate an integral of this kind correctly, collective variables λ_i are formally defined as functionals of an arbitrary configuration of φ: $\lambda_i = f_i\{\varphi\}$. For example, the center x_0 of an instanton can be defined by the relation

$$\int d^d x \varphi^4(x)(x - x_0) = 0, \qquad (4.1.13)$$

i.e.,

$$x_0 = \frac{\int d^d x \varphi^4(x) x}{\int d^d x \varphi^4(x)}. \qquad (4.1.14)$$

The integral in collective variables is performed by inserting the following partition of unity into the integrand in (4.1.11):

$$1 = \prod_{i=1}^{r} \int d\lambda_i \delta(\lambda_i - f_i\{\varphi\}), \qquad (4.1.15)$$

where $f_i\{\varphi\}$ can be defined as homogeneous functionals of φ of degree zero[5] (cf. (4.1.14)). If the arguments of the delta functions in (4.1.15) are linearized in the neighborhood of a saddle-point configuration,

$$\begin{aligned} 1 &= \prod_{i=1}^{r} \int d\lambda_i \delta(\lambda_i - f_i\{\varphi_c\} - (f_i'\{\varphi_c\}, \delta\varphi)) \\ &= \prod_{i=1}^{r} \int d\lambda_i \delta(\lambda_i - f_i\{\phi_c\} - \sqrt{g_c}(f_i'\{\phi_c\}, \delta\varphi)), \end{aligned} \qquad (4.1.16)$$

and the instanton is defined so that $\lambda_i - f_i\{\phi_c\} = 0$ (e.g., using a solution that is symmetric about the point $x = x_0$ in (4.1.14)), then ϕ_c is a function of λ_i, i.e., $\phi_c \equiv \phi_\lambda$. Perform a linear change of variables $\delta\varphi \longrightarrow \hat{S}\delta\varphi$ with $\det\hat{S} = 1$ to diagonalize the matrix $S''\{\phi_c\}$ and set

$$D\varphi = D'\varphi \prod_{i=1}^{r} d\tilde{\varphi}_i, \qquad (4.1.17)$$

where the r variables (denoted by the tilde) that correspond to the zero eigenvalues of $S''\{\phi_c\}$ and actually do not contribute to the exponential in (4.1.11) are factored out. Substituting (4.1.16) and (4.1.17) into (4.1.11), removing the delta functions by performing integration in $\delta\tilde{\varphi}_i$, and calculating the integral in $D'\varphi$, we obtain

$$\begin{aligned} Z_N &= cS_0^{-N}\Gamma\left(N + \frac{M + r}{2}\right), \\ S_0 &= S\{\phi_c\}, \end{aligned} \qquad (4.1.18)$$

$$\begin{aligned} c &= \frac{S_0^{-(M+r)/2}}{(2\pi)^{1+r/2}} \sqrt{-\frac{\det S''\{0\}}{\det[S''\{\phi_c\}]_{P'}}} \frac{1}{\det[f'\{\phi_c\}]_P} \\ &\times \int \prod_{i=1}^{r} d\lambda_i \phi_\lambda^{(1)} \ldots \phi_\lambda^{(M)}, \end{aligned} \qquad (4.1.19)$$

where $f'\{\phi_c\}$ is the operator defined by the matrix consisting of the columns $f_i'\{\phi_c\}$, and the subscripts P and P' denote projections onto the subspace spanned by the zero modes and its complement, respectively.[6] The ultraviolet divergences that arise when the constant c is calculated are eliminated by conventional renormalization of mass and charge. A general renormalization scheme of this kind was developed by Brezin and Parisi (see [8, 34]). Specific values of the parameters in (4.1.18) can be found in [7, 14, 34, 35], and the most general formal results were presented in [36–38].

According to (4.1.18), each degree of freedom associated with a zero mode contributes 1/2 to the argument of the gamma function. This resembles the classical equidistribution principle. A more careful analysis reveals a direct analogy. Indeed, the conventional partition function Z is a configuration-space integral of $\exp(-H/T)$. As the number r_{osc} of oscillatory degrees of freedom increases by unity, Z changes to $ZT^{1/2}$ and a corresponding 1/2 is added to specific heat [39]. Integral (4.1.4) is dominated by the exponential $\exp(-S\{\phi\}/g)$, and the coupling constant g plays the role of temperature. An increase by unity in the number r of zero modes corresponds to a decrease by unity in r_{osc} and change from Z to $Zg^{-1/2}$. To calculate the Lipatov asymptotic form, the factor $g^{-1/2}$ is estimated at the saddle point $g_c \sim 1/N$ (see (4.1.8)), Z_N is replaced by

---

[5] The result is actually independent of the particular form of the functionals [33], and only their linear independence is essential.

[6] In some cases, when $\det[f'\{\phi_c\}]_P$ depends on collective variables, it must be factored into the integral in $d\lambda_i$.

$Z_N N^{1/2}$, and 1/2 is added to the argument of the gamma function. In φ^4 theory with $d < 4$, the total number of zero modes is $r = d + n - 1$, including d instanton translations and $n - 1$ instanton rotations in a vector space. In a four-dimensional massless theory, there also exists a dilatation mode associated with scale invariance under variation of the instanton radius.

The equidistribution principle may be violated in the presence of soft modes associated with approximate symmetries: some degrees of freedom resemble zero modes in the first approximation (see Fig. 4a), but a more accurate analysis shows that they correspond to motions in a slowly varying potential (Fig. 4b), which may have a nonanalytic minimum (Fig. 4c). The contribution of such a mode to the argument of a delta function is neither zero nor 1/2.

One problem arising in the presence of soft modes is that the instanton ϕ_c is only an approximate solution to the equation $S'\{\phi\} = 0$, while there may not exist any exact solution at all. Accordingly, the linear terms in the expansion of ϕ_c in powers of $\delta\phi$ must be accurately eliminated. The collective variable characterizing the location of a particle moving in a slowly varying potential (see Fig. 4b) can be formally defined as a functional on an arbitrary instanton configuration: $z = f\{\phi\}$. An extremum of the action is sought under the constraint $f\{\phi\}$ = const (i.e., for a constant z) and then an integral in z is calculated. Then, the instanton is determined by the equation

$$S'\{\phi_c\} - \mu f'\{\phi_c\} = 0, \tag{4.1.20}$$

where μ is a Lagrange multiplier, and the integration with respect to z is performed by inserting the following partition of unity into the functional integral:

$$\begin{aligned} 1 &= \int dz \delta(z - f\{\phi\}) \\ &= \int dz \delta(z - f\{\phi_c\} - (f'\{\phi_c\}, \delta\phi)). \end{aligned} \tag{4.1.21}$$

Using the condition $z = f\{\phi_c\}$ to define an instanton, one obtains

$$\begin{aligned} Z(g) &= \int D\varphi \varphi^{(1)} \ldots \varphi^{(M)} \\ &\times \exp\left\{-\frac{S\{\phi_c\} + (S'\{\phi_c\}, \delta\phi) + \frac{1}{2}(\delta\phi, S''\{\phi_c\}\delta\phi)}{g}\right\} \\ &\times \int dz \delta(-(f'\{\phi_c\}, \delta\phi)), \end{aligned} \tag{4.1.22}$$

and the linear in $\delta\phi$ terms in the exponential are canceled by a delta function by virtue of condition (4.1.20). Since ϕ_c is a function of z, integration with respect to $D\varphi$ results in a nontrivial integral in z, which corresponds to the motion in a slowly varying potential

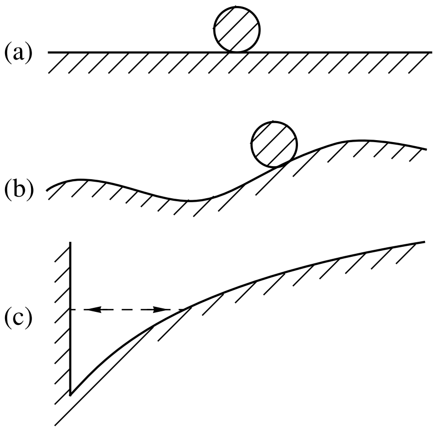


Fig. 4. Soft mode: (a) zero mode (in the first approximation); (b) motion in a slowly varying potential (in a more detailed analysis); (c) the potential may have a nonanalytic minimum.

(Fig. 4b). Note that the rearrangements performed in (4.1.20)–(4.1.22) are not restrictive, and any degree of freedom can be treated as a soft mode. However, this should be done only if the validity of the Gaussian approximation is questionable.

Examples of soft modes are the dilatations in the massive four-dimensional or $(4 - \epsilon)$-dimensional φ^4 theory [37, 38] and the variation of the distance between elementary instantons in a two-instanton configuration (see Sections 4.4 and 9). The analysis above shows that the shift b in the argument of the gamma function includes contributions of external lines ($M/2$) and zero modes ($r/2$), as well as the contribution ν due to soft modes.

4.2. Quantum Electrodynamics

The simplest functional integral in quantum electrodynamics (vacuum integral) has the form

$$\begin{aligned} Z = \int DAD\bar{\psi}D\psi \exp\Bigg\{-\int d^4x \Bigg[\frac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)^2 \\ + \bar{\psi}(i\gamma_\nu\partial_\nu - m + e\gamma_\nu A_\nu)\psi\Bigg]\Bigg\}, \end{aligned} \tag{4.2.1}$$

where A_ν is the vector potential, and $\bar{\psi}$ and ψ denote fermionic fields represented in terms of Grassmann variables. These are abstract quantities for which formal algebraic operations are defined [40, 41], and the standard Lipatov method cannot be applied directly when they contribute to an exponential. A remedy can be found by noting that the action is quadratic in the fermionic fields and the Gaussian integral is easily calculated in a Grassmann algebra as the determinant of the corresponding quadratic form (see [15]):

$$Z = \int DA\det(i\gamma_\nu\partial_\nu - m + e\gamma_\nu A_\nu) \times \exp\left\{-\frac{1}{4}\int d^4x(\partial_\mu A_\nu - \partial_\nu A_\mu)^2\right\}. \qquad (4.2.2)$$

If det(…) is represented as exp{log det(…)}, then the resulting effective action contains only the vector potential A_ν, which can be treated by Lipatov's method.

The determinant of an operator is too difficult to be used constructively, and considerable effort has been applied to reduce it to a tractable form (see [22, 23]). In particular, it was difficult to establish the general properties of a saddle-point configuration when no tractable expression for effective action was available [22]. It was found that the saddle-point value of $eA_\nu(x)$ is large. Accordingly, one can make use of the asymptotic form of the determinant as $e \longrightarrow i\infty$ since the fastest growth corresponds to a pure imaginary e [20]:

$$\det(i\gamma_\nu\partial_\nu - m + e\gamma_\nu A_\nu) = \exp\left\{\frac{e^4}{12\pi^2}\int d^4x(A_\nu^2)^2\right\}. \qquad (4.2.3)$$

Expression (4.2.3) is not gauge invariant. It is valid only in a restricted set of gauges for which the scale of vector-potential variation is comparable to that of the physical electromagnetic field, which is treated as semiclassical.[7] Actually, these gauges are close to the Lorentz gauge, as can be shown by considering configurations characterized by high symmetry [20, 23].

Substituting (4.2.3) into (4.2.2), one obtains the functional integral containing the effective action

$$S_{\text{eff}}\{A\} = \int d^4x\left\{\frac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)^2 - \frac{4}{3}g^2(A_\nu^2)^2\right\}, \qquad (4.2.4)$$

for which asymptotic results of perturbation theory can be found in the saddle-point approximation. The structure of these asymptotics is determined by the homogeneity properties of the action, which are analogous to those in φ^4 theory with g^2 used as a coupling constant. According to Section 4.1, the general term of the asymptotics has the form $cS_0^{-N}\,\Gamma(N + b)g^{2N}$, where S_0 is an instanton action. Actually, the series expansion is developed in terms of arbitrary (not only even) powers of g, and the substitution $N \longrightarrow N/2$ leads to $cS_0^{-N/2}\,\Gamma(N/2 + b)g^N$ as the Nth-order contribution. This formal substitution is valid, because the direct expansion of (4.2.4) in terms of the last summand is not correct, because the functional integration would involve configurations for which (4.2.4) is not valid. Calculation must be performed by the saddle-point method, which yields a continuous function of N, and the fact that its values must be taken at integer or half-integer points is an external condition.

By using the value of the instanton action, the coefficients of high-order terms in the expansion of (4.2.1) are expressed as follows [14, 15]:

$$Z_N = \text{const}\, S_0^{-N/2}\Gamma\left(\frac{N+r}{2}\right), \quad S_0 = \frac{4\pi^3}{3^{3/2}}. \qquad (4.2.5)$$

Here, the total number r of zero modes is 11, including four translations, a scale transformation, and six four-dimensional rotations, since the symmetry of an instanton is similar to that of an irregular solid.

The scheme developed above can be applied to calculate other quantities [42]. The most general vertex in QED contains M photon integrals and $2L$ electron lines and corresponds to the functional integral

$$Z_{M,L} = \int DAD\bar{\psi}D\psi A(x_1) \ldots A(x_M)\psi(y_1)\bar{\psi}(z_1)\ldots\psi(y_L)\bar{\psi}(z_L) \times \exp\left\{-\int d^4x\left[\frac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)^2 + \bar{\psi}(i\gamma_\nu\partial_\nu - m + e\gamma_\nu A_\nu)\psi\right]\right\}. \qquad (4.2.6)$$

Integration over the fermionic fields results in

$$Z_{M,L} = \int DAA(x_1) \ldots A(x_M)G(y_1, z_1)\ldots G(y_L, z_L) \times \det(i\gamma_\nu\partial_\nu - m + e\gamma_\nu A_\nu) \times \exp\left\{-\frac{1}{4}\int d^4x(\partial_\mu A_\nu - \partial_\nu A_\mu)^2\right\} + \ldots. \qquad (4.2.7)$$

Here, $G(x, x')$ is the Green function of the Dirac operator,

$$(i\gamma_\nu\partial_\nu - m + e\gamma_\nu A_\nu)G(x, x') = \delta(x - x'), \qquad (4.2.8)$$

and the ellipsis stands for terms with different pairings of $\psi(y_i)$ and $\bar{\psi}(z_k)$. The structure of the result can be found by performing calculations as demonstrated above, i.e., essentially by dimensional analysis. It can readily be shown that $e_c \sim N^{-1/4}$ and $A_c(x) \sim N^{1/2}$ for a saddle-point configuration. To determine the dimension

---

[7] The general scheme for deriving expressions analogous to (4.2.3) is illustrated in Section 4.3 by using a simpler example.

of $G(x, x')$, write out the Dyson equation that follows from (4.2.8):

$$\begin{aligned} G(x, x') &= G_0(x-x') \\ &- \int d^4 y G_0(x-y) e\gamma_\nu A_\nu(y) G(y, x'). \end{aligned} \tag{4.2.9}$$

To elucidate the structure of the solution, consider the scalar counterpart of (4.2.9) and assume that the function $A_\nu(x)$ is localized within a small neighborhood of $x = 0$. Then, the equation is easily solved by setting $G(y, x') \approx G(0, x')$:

$$\begin{aligned} G(x, x') &= G_0(x-x') \\ &- \frac{G_0(-x')\int d^4 y G_0(x-y) e\gamma_\nu A_\nu(y)}{1+\int d^4 y G_0(-y) e\gamma_\nu A_\nu(y)}. \end{aligned} \tag{4.2.10}$$

Since $eA_\nu(x) \sim N^{1/4}$ and (4.2.10) tends to a finite limit as $e \longrightarrow \infty$, the result is $G(x, x') \sim N^0$. It is reasonable to expect that its validity is independent of the assumptions used in its derivation. The Nth-order contribution to integral (4.2.6) has the form

$$\mathrm{const}\left(\frac{3^{3/2}}{4\pi^3}\right)^{N/2}\Gamma\left(\frac{N+r+M}{2}\right)(-g)^N \tag{4.2.11}$$

for even M and a similar form multiplied by $eN^{1/4}$ for odd M.

4.3. Other Fermionic Models

As an additional example of fermionic model, consider the Yukawa interaction [15], to which description of electron–phonon interaction in metals can be reduced with minor changes:

$$\begin{aligned} Z = \int D\varphi D\bar{\psi} D\psi \exp\bigg\{-\int d^d x\bigg[\frac{1}{2}(\partial_\mu\varphi)^2 \\ + \frac{1}{2}m^2\varphi^2 + \bar{\psi}(i\gamma_\nu\partial_\nu + M)\psi + \lambda\bar{\psi}\varphi\psi\bigg]\bigg\}. \end{aligned} \tag{4.3.1}$$

Integration over the fermionic fields results in

$$\begin{aligned} Z &= \int D\varphi \det(i\gamma_\nu\partial_\nu + M + \lambda\varphi) \\ &\times \exp\left\{-\int d^d x\left[\frac{1}{2}(\partial_\mu\varphi)^2 + \frac{1}{2}m^2\varphi^2\right]\right\}. \end{aligned} \tag{4.3.2}$$

The transformation of the fermion determinant begins with the solution of an analogous problem for the determinant of the Schrödinger operator normalized to the determinant of the unperturbed problem (normalization of this kind is due to the normalization of a functional integral to the vacuum integral in the interaction-free theory):

$$\begin{aligned} D(z) &= \frac{\det[-\Delta - E + zV(x)]}{\det[-\Delta - E]} \\ &= \det\left[1 + z\frac{V(x)}{-\Delta - E}\right]. \end{aligned} \tag{4.3.3}$$

It can easily be shown that

$$D(z) = \prod_s\left(1 + \frac{z}{\mu_s}\right), \tag{4.3.4}$$

where μ_s denotes the eigenvalues of the problem

$$\{-\Delta - E - \mu_s V(x)\}e_s(x) = 0. \tag{4.3.5}$$

The number s of energy states below E for an electron moving in a semiclassical potential $-\mu V(x)$ can be found by invoking the Thomas–Fermi model. By using the local Fermi momentum

$$p(x) = \sqrt{E + \mu V(x)} \approx [\mu V(x)]^{1/2}, \tag{4.3.6}$$

it is expressed as

$$\begin{aligned} s &= \int n(x) d^d x = \int \frac{K_d}{d} p^d(x) d^d x \\ &\approx \int \frac{K_d}{d}[\mu V(x)]^{d/2} d^d x, \end{aligned} \tag{4.3.7}$$

where $n(x)$ is the local electron density and K_d is the area of a d-dimensional unit sphere divided by $(2\pi)^d$. Since the value of μ_s in (4.3.5) corresponds to the condition that exactly s electron energy states lie below E, expression (4.3.7) describes the asymptotic behavior of μ_s for large s:

$$s = A\mu_s^{d/2}, \quad A = \frac{K_d}{d}\int [V(x)]^{d/2} d^d x. \tag{4.3.8}$$

Now, the value of (4.3.4) can easily be estimated at large z:

$$\begin{aligned} \ln D(z) &= \sum_{s\sim 1}^{\infty}\ln\left(1 + \frac{z}{\mu_s}\right) \\ &\approx A\frac{d}{2}\int\limits_{\sim A^{-2/d}}^{\infty} d\mu_s \mu_s^{d/2-1}\ln\left(1 + \frac{z}{\mu_s}\right) \\ \approx A\frac{d}{2}z^{d/2}&\int\limits_0^{\sim zA^{2/d}}\frac{dx}{x^{1+d/2}}\ln(1+x) \approx Az^{d/2}\frac{\pi}{\sin(\pi d/2)}. \end{aligned} \tag{4.3.9}$$

When $d < 2$, the upper limit in the integral can be set at infinity, and the integral is calculated by changing from x to ax in the logarithm and differentiating the result

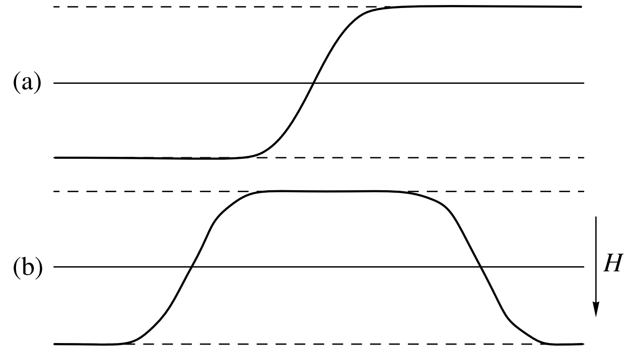


Fig. 5. Domain wall as example of topological instanton: (a) degenerate vacuum; (b) vacuum state is split into an instanton–anti-instanton pair corresponding to a saddle-point configuration.

with respect to a. The substitution of $K_d = 2^{1-d}\pi^{-d/2}/\Gamma/(d/2)$ yields the final result

$$D(z) = \exp\left\{-\frac{\Gamma(-d/2)}{(4\pi)^{d/2}}\int d^d x[zV(x)]^{d/2}\right\}. \quad (4.3.10)$$

It is well known that the Dirac operator can be obtained by factorization of the Klein–Gordon operator transformed into the following Schrödinger operator by Wick rotation:

$$-\Delta + m^2 = (i\gamma_\nu\partial_\nu + m)(-i\gamma_\nu\partial_\nu + m). \quad (4.3.11)$$

An analogous relation holds after m is replaced with $\lambda\varphi(x)$, where $\varphi(x)$ is a slowly varying function. For large λ, it can readily be shown that

$$\det\left(\frac{i\gamma_\nu\partial_\nu + M + \lambda\varphi}{i\gamma_\nu\partial_\nu + M}\right) \approx \sqrt{\det\left(\frac{-\Delta + \lambda^2\varphi^2(x)}{-\Delta}\right)} \approx \exp\left\{-\frac{\Gamma(-d/2)}{2(4\pi)^{d/2}}\int d^d x[\lambda\varphi(x)]^d\right\} \quad (4.3.12)$$

and (4.3.2) yields the effective action

$$S_{\text{eff}} = \int d^d x\left[\frac{1}{2}(\partial_\mu\varphi)^2 + \frac{1}{2}m^2\varphi^2 + \frac{\Gamma(-d/2)}{2(4\pi)^{d/2}}g^{d/2}\varphi^d(x)\right], \quad (4.3.13)$$

where $g = \lambda^2$ is used as an effective coupling constant in an expansion of integral (4.3.1).

When $2 \le d < 4$ or $d \ge 4$, the divergence of the integral in (4.3.9) is eliminated by renormalizing both mass and charge [34] so that $\ln(1 + x)$ is replaced with $\ln(1 + x) - x$ and $\ln(1 + x) - x + x^2/2$, respectively. The integral is calculated by changing from x to ax and differentiating the result with respect to a. The result obtained for $d \ne 2, 4$ is formally identical to (4.3.13). If $d = 2$ or 4, then the calculation can be performed to logarithmic accuracy by taking into account the finite upper limit in the last integral in (4.3.9).

Effective action (4.3.13) can be rewritten as

$$S_{\text{eff}}\{g, \varphi\} = \frac{S\{\varphi\}}{g^{1/\alpha}}, \quad (4.3.14)$$

where

$$\phi = \varphi g^{1/2\alpha}, \quad \alpha = \frac{d-2}{d},$$

and the change to $\tilde{g} = g^{1/\alpha}$ and $\tilde{N} = \alpha N$ is performed to represent the integral expression for the expansion coefficients in a form similar to that in φ^4 theory. The result is similar to φ^4 theory up to the change $N \longrightarrow \alpha N$:

$$Z_N = cS_0^{-\alpha N}\Gamma(\alpha N + b). \quad (4.3.15)$$

4.4. Degenerate Vacuum

Calculations of the Lipatov asymptotic forms for theories with degenerate vacuum require special analysis.

The simplest example is the one-dimensional Ising ferromagnet with a doubly degenerate ground state with all spins either up or down. In addition to these vacuums, there exists a classical domain-wall solution (an example of *topological instanton*), which corresponds to transition between the two degenerate vacuums (Fig. 5a). The issues to be resolved in this model arise from the following: (a) the contributions of topological instantons to asymptotic expressions for expansion coefficients are strictly imaginary, which implies that they are insignificant in some sense; (b) generally, the absence of other nontrivial classical solutions to problems of this kind is established by special theorems.

To elucidate these issues, suppose that the degeneracy of the vacuum states is eliminated by applying a magnetic field aligned with the ferromagnet's axis. Then, domain-wall-like excitations cannot exist, because they are associated with an infinitely large additional energy (in the infinite-volume limit). However, they can exist in the form of instanton–anti-instanton pairs (Fig. 5b). If the interaction between the components of such a pair is repulsive, then there exists a stationary state characterized by an average distance between the components. As the distance decreases with increasing magnetic field strength, a localized instanton of the type considered in Lipatov's method is obtained. As the distance increases with decreasing magnetic field strength, the amplitude of distance fluctuations increases, and the pair breaks up into free instanton and anti-instanton in the limit of strictly degenerate vacuum.

Because of strong fluctuations of the distance between the components, the saddle-point approximation can be applied in Lipatov's method only if $N \gg 1/\epsilon$ (rather than $N \gg 1$), where ϵ is a small parameter characterizing the difference between the vacuums. However, there exists an intermediate asymptotic regime that is virtually independent of ϵ for $1 \ll N \ll 1/\epsilon$. As $\epsilon \longrightarrow 0$, this intermediate asymptotic behavior approaches the true asymptotics of the degenerate problem. This implies that the latter asymptotics can be found without analyzing the case of $\epsilon = 0$. For finite ϵ, single instantons do not exist and therefore do not contribute to the asymptotics of perturbation theory, which are determined by the instanton–anti-instanton pair.

The physical interpretation outlined here was developed in an analysis of the quantum mechanical double-well potential problem [43]. Analogous interpretations can be expected to be valid for other theories with degenerate vacuum, among which Yang–Mills theories are of particular interest.

4.5. Yang–Mills Theory and QCD

The topological instanton found in [44] for the Yang–Mills theory was the earliest evidence of the existence of degenerate vacuum in QCD. In [45], the saddle-point calculation of a functional integral was performed for the one-instanton configuration of $SU(2)$ Yang–Mills fields coupled to fermions and scalar particles. This result was extended to arbitrary $SU(N_c)$ symmetry in [46]. In an analysis of saddle-point configurations performed for the Yang–Mills field coupled to a scalar field in [18], a continuous transformation of the φ^4 theory instanton into a saddle-point configuration for the pure Yang–Mills theory was found. The latter configuration was shown to correspond to an instanton–anti-instanton pair; i.e., a physical interpretation analogous to that discussed above was developed. The result obtained in [45] for a single instanton was used in [17] to calculate the contribution of the instanton–anti-instanton configuration to asymptotic behavior in perturbation theory for $SU(2)$ Yang–Mills fields. The Lipatov asymptotic forms for realistic QCD were calculated in [24–26]. A general scheme of these calculations is presented below.

As a first step, a rule for combining instantons [33] is formulated for the functional integral

$$Z_M(g) = \int DAA^{(1)}A^{(2)}\dots A^{(M)}\exp(-S\{A, g\}), \quad (4.5.1)$$

where $A^{(i)}$ is a bosonic field, and the superscript i stands for both coordinate and internal degrees of freedom. Suppose that the action $S\{A, g\}$ is rewritten as $S\{B\}/g^2$ by changing from A to B/g and the equation $S'\{B\} = 0$ has an instanton solution B_c. By following the scheme developed in Section 4.1, it can readily be shown that the one-instanton contribution to $Z_M(g)$ has the form

$$\begin{aligned} Z_M^{(1)}(g) &= c_0 g^{-M-r} e^{-S_0/g^2} \\ &\times \int \prod_{i=1}^{r} d\lambda_i B_\lambda^{(1)} B_\lambda^{(2)} \dots B_\lambda^{(M)}, \end{aligned} \quad (4.5.2)$$

where $S_0 = S\{B_c\}$, r is the number of zero modes, λ_i denotes the corresponding collective variables, and B_λ is the instanton configuration depending on these variables.

If B_c is the combination $B_\lambda + B_{\lambda'}$ of elementary instantons, then the corresponding two-instanton contribution can be represented as the sum of terms of the form

$$\begin{aligned} Z_{LL'}(g) &= c_0^2 g^{-M-2r} e^{-2S_0/g^2} \\ &\times \int \prod_{i=1}^{r} d\lambda_i d\lambda'_i B_\lambda^{(1)} \dots B_\lambda^{(L)} B_{\lambda'}^{(1)} \dots B_{\lambda'}^{(L')} \\ &\times \exp\left(-\frac{S_{\text{int}}(B_\lambda, B_{\lambda'})}{g^2}\right) \end{aligned} \quad (4.5.3)$$

with $L + L' = M$. The instanton–instanton interaction $S_{\text{int}}(B_\lambda, B_{\lambda'})$ is defined by the relation

$$\begin{aligned} &S\{B_\lambda + B_{\lambda'}\} \\ &\equiv S\{B_\lambda\} + S\{B_{\lambda'}\} + S_{\text{int}}(B_\lambda, B_{\lambda'}). \end{aligned} \quad (4.5.4)$$

When the interaction is neglected, the right-hand side of (4.5.3) reduces to the product of two expressions having the form of (4.5.2), with $M = L$ and $M = L'$. Due to the exponential factor, the instanton–instanton interaction is limited by the condition $S_{\text{int}}(B_\lambda, B_{\lambda'}) \lesssim g^2$. When g is small, this condition is insignificant, and the overlap of B_λ and $B_{\lambda'}$ can be neglected. The resulting sum in L and L' contains only the terms with $L = M$, $L' = 0$ and $L = 0$, $L' = M$, which are obviously equal. The ensuing factor 2 is canceled by the combinatorial factor $1/2!$ introduced to preclude double counting of configurations. The resulting two-instanton contribution,

$$\begin{aligned} Z_M^{(2)}(g) &= c_0^2 g^{-M-2r} e^{-2S_0/g^2} \\ &\times \int \prod_{i=1}^{r} d\lambda_i d\lambda'_i B_\lambda^{(1)} \dots B_\lambda^{(M)} \exp\left(-\frac{S_{\text{int}}(B_\lambda, B_{\lambda'})}{g^2}\right), \end{aligned} \quad (4.5.5)$$

entails a rule for combining instantons: in addition to the information contained in (4.5.2), it is necessary to know the instanton–instanton interaction in the domain where the interaction is weak.

In relativistic scale-invariant theories, the integrals with respect to the radius ρ and center x_0 of an instanton in (4.5.2) can be performed separately [41], while λ_i are treated as internal degrees of freedom:

$$\begin{aligned} Z_M^{(1)}(g) &= c_H g^{-M-r} e^{-S_0/g^2} \int \prod_i d\lambda_i \int d^4 x_0 \\ &\times \int d\rho \rho^{-M-5} e^{\nu \ln \mu\rho} B_\lambda(y_1) \ldots B_\lambda(y_M), \end{aligned} \tag{4.5.6}$$

where $y_i = (x_i - x_0)/\rho$, $\nu = -\beta_2 S_0$, β_2 is the lowest order nonvanishing expansion coefficient for the Gell-Mann–Low, μ is the value of momentum at the normalization point, and $\exp(\nu \ln \mu\rho)$ is uniquely factored out by virtue of the renormalizability condition [41]. Formula (4.5.6) is consistent with 't Hooft's result for $SU(N_c)$ Yang–Mills fields ($S_0 = 8\pi^2$, $r = 4N_c$, $\nu = (11N_c - 2N_f)/3$) [45, 46] and (with g^2 replaced by g) with the corresponding result in φ^4 theory [7, 14, 37].

The QCD Lagrangian has the form

$$\begin{aligned} L &= -\frac{1}{4}(F^a_{\mu\nu})^2 - \frac{1}{2\alpha}(\partial_\mu A^a_\mu)^2 \\ &+ \sum_f \bar{\psi}_f \hat{D} \psi_f + \partial_\mu \bar{\omega}^a (\partial_\mu \omega^a - \bar{g} f^{abc} \omega^b A^c_\mu), \\ F^a_{\mu\nu} &= \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + \bar{g} f^{abc} A^b_\mu A^c_\nu, \\ \hat{D} &= i\gamma_\mu(\partial_\mu - i\bar{g} A^a_\mu T^a), \end{aligned} \tag{4.5.7}$$

where A^a_ν, ψ_f, and ω^a denote gluon, quark, and ghost fields, respectively; T^a and f^{abc} are the generators of the fundamental representation and structure constants of the Lie algebra, respectively; α is the gauge parameter; and the subscript "f" denotes the types of quarks, whose total number is N_f. The preexponential factor in the most general functional integral for QCD contains M gluon fields, $2L$ ghost fields, and $2K$ quark fields:

$$\begin{aligned} Z_{MLK} &= \int DAD\bar{\omega}D\omega D\bar{\psi}D\psi A(x_1) \\ &\ldots A(x_M)\omega(y_1)\bar{\omega}(\bar{y}_1)\ldots\omega(y_L)\bar{\omega}(\bar{y}_L) \\ &\times \psi(z_1)\bar{\psi}(\bar{z}_1)\ldots\psi(z_K)\bar{\psi}(\bar{z}_K) \\ &\times \exp(-S\{A, \bar{\omega}, \omega, \bar{\psi}, \psi\}), \end{aligned} \tag{4.5.8}$$

where the vector indices that are not essential for the present analysis are omitted. By replacing A with $B/\bar{g}$, the Euclidean action is rewritten as

$$\begin{aligned} S\{A, \bar{\omega}, \omega, \bar{\psi}, \psi\} &\longrightarrow \frac{S\{B\}}{\bar{g}^2} \\ &+ \int d^4 x \left[\bar{\omega} \hat{Q} \omega + \sum_f \bar{\psi}_f \hat{D} \psi_f \right]. \end{aligned} \tag{4.5.9}$$

Integration over the fermionic fields results in

$$\begin{aligned} Z_{MLK} &= (1/\bar{g})^M \int DAB(x_1) \\ &\ldots B(x_M) G(y_1, \bar{y}_1) \ldots G(y_L, \bar{y}_L) \\ &\times \tilde{G}(z_1, \bar{z}_1) \ldots \tilde{G}(z_K, \bar{z}_K) \det\hat{Q} (\det\hat{D})^{N_f} \\ &\times \exp\{-S\{B\}/\bar{g}^2\} + \ldots, \end{aligned} \tag{4.5.10}$$

where G and $\tilde{G}$ are the Green functions of the operators $\hat{Q}$ and $\hat{D}$, and the ellipsis stands for terms with different pairings. It is important here that $S\{B\}$, G, and $\tilde{G}$ are independent of $\bar{g}$. Functional integral (4.5.10) is dominated by the Yang–Mills action, and the corresponding one-instanton contribution can be written out by analogy with (4.5.2). The asymptotic behavior in perturbation theory is determined by an instanton–anti-instanton contribution calculated by analogy with (4.5.5). The instanton–instanton interaction is specified by introducing a conformal parameter ξ:

$$S_{\text{int}} = -h\xi, \quad \xi = \frac{\rho_I^2 \rho_A^2}{(R^2 + \rho_I^2 + \rho_A^2)^2}, \tag{4.5.11}$$

where ρ_I and ρ_A denote the instanton and anti-instanton radii, R is the distance between their centers, and $h = h(\lambda, \lambda')$ depends on their mutual orientation in the isotopic space [24]. Next, it should be noted that $\det\hat{D}\{B_\lambda + B_{\lambda'}\} \neq \det\hat{D}\{B_\lambda\}\det\hat{D}\{B_{\lambda'}\}$ since $\det\hat{D}\{B_\lambda\}$ does not vanish only if the finite quark mass is taken into account, whereas $\det\hat{D}\{B_\lambda + B_{\lambda'}\}$ is determined by the instanton–instanton interaction and is finite in the massless limit (see [24]),

$$\det\hat{D}\{B_\lambda + B_{\lambda'}\} = \text{const}\,\xi^{3/2}. \tag{4.5.12}$$

By factoring the integrals in the instanton radii and centers and changing to the momentum representation, the instanton–anti-instanton contribution is rewritten as

$$\begin{aligned} Z^{(IA)}_{MLL'} &= \frac{\text{const}}{\bar{g}^{M+2r}} e^{-2S_0/g^2} \int \prod_i d\lambda_i d\lambda'_i \\ &\times \int d\rho \rho^{3M + 6L + 5L' - 5} e^{2\nu \ln \mu\rho} \langle B_\lambda \rangle_{\rho p_1} \ldots \langle B_\lambda \rangle_{\rho p_M} \\ &\times \langle G_\lambda \rangle_{\rho k_1, \rho k'_1} \ldots \langle G_\lambda \rangle_{\rho k_L, \rho k'_L} \langle \tilde{G}_\lambda \rangle_{\rho q_1, \rho q'_1} \\ &\ldots \langle \tilde{G}_\lambda \rangle_{\rho q_{L'}, \rho q'_L} \int_0^{\sim 1} \frac{d\xi}{\xi^{1 + \nu/2 - 3N_f/2}} e^{-h(\lambda, \lambda')\xi/\bar{g}^2} + \ldots, \end{aligned} \tag{4.5.13}$$

where $\langle B \rangle_k$, $\langle G \rangle_{k, k'}$, and $\langle \tilde{G} \rangle_{k, k'}$ denote Fourier components of $B(x)$, $G(x, x')$, and $\tilde{G}(x, x')$, respectively; $\rho \equiv$

ρ_{I}; use is made of the fact that the dominant contribution is due to the region where $R \sim \rho_{\mathrm{A}} \gg \rho_{\mathrm{I}}$; and relation (18) from [24] is taken into account.

By following [17] (see also [33]), the last integral is replaced by the corresponding jump across the cut. Then, (4.5.13) yields a jump in the total value of $Z_{MLL'}$ across the cut:

$$\Delta Z_{MLL'}(\bar{g}) = i\,\mathrm{const}\left(\frac{1}{\bar{g}}\right)^{M+2r+\nu-3N_{\mathrm{f}}} \times \exp\left(-\frac{2S_0}{\bar{g}^2}\right), \tag{4.5.14}$$

where the independence of B_λ, G_λ, and $\tilde{G}_\lambda$ of $\bar{g}$ is used and all integrals are assumed to be convergent.[8] By virtue of relation (3.17) between the jump across the cut and the asymptotic form of expansion coefficients, the Nth-order contribution to $Z_{MLL'}$ is

$$[Z_{MLK}]_N \bar{g}^{2N} = \mathrm{const}(16\pi^2)^{-N} \times \Gamma\left(N + \frac{M}{2} + 4N_{\mathrm{c}} + \frac{11(N_{\mathrm{c}} - N_{\mathrm{f}})}{6}\right)\bar{g}^{2N} \tag{4.5.15}$$

for even M and by a similar form multiplied by $\bar{g}N^{1/2}$ for odd M [17, 24, 47]. This result is analogous to that discussed in Section 4.1: the term $M/2$ in the argument of the gamma function is determined by the number of external lines, $4N_{\mathrm{c}}$ is half the number of zero modes, and $11(N_{\mathrm{c}} - N_{\mathrm{f}})/6$ is the additional contribution of the soft mode corresponding to variation of the instanton–anti-instanton distance. Specific values of the constant factor were calculated in [17, 24, 25].

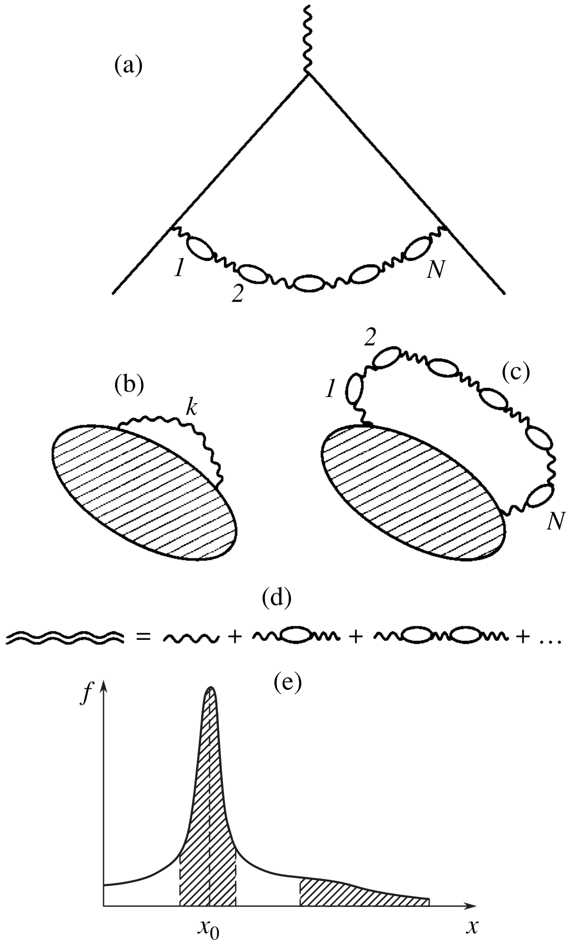


Fig. 6. (a) Example of single QED diagram of Nth-order giving contribution N! [49]. Renormalon diagram: (b) an internal photon line is singled out; (c) a chain of electron "bubbles" is inserted; (d) insertions into the photon line correspond to interaction "dressing." (e) Saddle-point method is formally applicable, but leads to incorrect results.

5. RENORMALONS AND MATHEMATICAL SUBSTANTIATION OF LIPATOV'S METHOD

5.1. 't Hooft's Argumentation

The invention of Lipatov's method was widely recognized, and it was immediately applied to almost every topical problem in theoretical physics (see [19]). However, the validity if Lipatov's method was questioned as early as in 1977. The criticism dates back to [49], where the following interesting remark was made. Lipatov's result (2.5) is obviously associated with the factorially increasing number of diagrams of order $(ag)^N$. However, this interpretation is incorrect in the general case: in particular, there exist Nth-order diagrams (with long chains of "bubbles") whose contributions are proportional to N! (see Fig. 6a), which were called renormalons since they arise only in renormalizable theories:[9] even though the example discussed in [49] and illustrated by Fig. 6a was taken from quantum electrodynamics, analogous diagrams arise in QCD and four-dimensional φ^4 theory. Strictly speaking, Lautrup's remark is inconsequential, since Lipatov's method relies on formal calculation of functional integral (2.4) and does not involve any statistical analysis of diagrams. Therefore, it should be expected that the renormalon contributions are subsumed under (2.5).

However, 't Hooft claimed in [50] that renormalons provide an independent mechanism of divergence of perturbation series and their contribution is not contained in the Lipatov asymptotics. The argumentation put forward by 't Hooft relies on an analysis of the analyticity properties of Borel transforms. Indeed, the

---

[8] In the quark–quark correlation function, the integral in ρ involves divergences. The method for eliminating them proposed in [24, 26] evokes doubts [48]. For $M \geq 1$, the integral is convergent.

[9] This is true for theories with running coupling, which involve logarithmic contributions. No renormalons arise in superrenormalizable theories.

Borel transform of the function represented by a series with expansion coefficients $ca^N\Gamma(N+b)$ has a singular point at $z = 1/a$:

$$B(z) = \sum_N ca^N N^{b-1} z^N \sim (1-az)^{-b}, \quad za \longrightarrow 1. \tag{5.1}$$

Thus, the value of a in (2.5) determines the location of a singular point in the Borel plane. This conclusion was obtained by 't Hooft without reference to Lipatov's method. Representing action as $S\{\phi\}/g$ (see Section 4), one can rewrite a general functional integral and the definition of Borel transform (3.12) as follows:

$$Z(g) = \int D\phi \exp\left(-\frac{S\{\phi\}}{g}\right), \tag{5.2}$$

$$Z(g) = \int_0^\infty dx e^{-x/g} B(x), \tag{5.3}$$

where the factors g^n are omitted since they cancel out when Green functions are calculated as ratios of two functional integrals. Then, the Borel transform of functional integral (5.2) is

$$B(z) = \int D\phi \delta(z - S\{\phi\}) = \oint_{z = S\{\phi\}} \frac{d\sigma}{|S'\{\phi\}|}, \tag{5.4}$$

where $|S'\{\phi\}|$ is the absolute value of the vector defined in (4.1.6), and the last integral is calculated over the hypersurface $z = S\{\phi\}$. If $\phi_c(x)$ is an instanton, then $S'\{\phi_c\} = 0$ and (5.4) has a singular point at $z = S\{\phi_c\}$, which coincides with $1/a$ for the instanton characterized by the least action S_0. Furthermore, there exist singular points at the points mS_0, which correspond to instantons at infinity, and singularities corresponding to instantons of other types. If $z = S_0$ is the singular point nearest to the origin, then Lipatov asymptotic form (2.5) is obtained. However, 't Hooft hypothesized that singularities other than instantons may exist, in which case the asymptotic behavior of expansion coefficients is determined by the non-instanton singular point nearest to the origin.

Renormalons were considered by 't Hooft as a possible new mechanism of singularity formation. The virtual photon line with momentum k in an arbitrary QED diagram (see Fig. 6b) represents an large-momentum integral of the form

$$\int d^4 k k^{-2n}, \tag{5.5}$$

where n is integer. After all renormalizations are performed, the integral is convergent and $n \geq 3$. When N electron bubbles are inserted into the photon line (see Fig. 6c), the integrand is multiplied by $\ln^N(k^2/m^2)$ (m is electron mass), and the resulting integral is proportional to $N!$. Insertions in the photon line correspond to "dressed" coupling. Accordingly, g_0 is replaced by a running coupling constant in the integrand of (5.5). Summation of diagrams of the form shown in Fig. 6c is equivalent to the use of the one-loop approximation $\beta(g) = \beta_2 g^2$ for the Gell-Mann–Low function and leads to a well-known result:

$$g(k^2) = \frac{g_0}{1 - \beta_2 g_0 \ln(k^2/m^2)}. \tag{5.6}$$

The integral over the region of $k^2 \gtrsim m^2$ yields

$$\int d^4 k k^{-2n} g(k^2) = g_0 \sum_N \int d^4 k k^{-2n} \left(\beta_2 g_0 \ln \frac{k^2}{m^2}\right)^N \sim g_0 \sum_N N! \left(\frac{\beta_2}{n-2}\right)^N g_0^N. \tag{5.7}$$

As a result of Borel summation, renormalon singularities are obtained at the points[10]

$$z_n = \frac{n-2}{\beta_2}, \quad n = 3, 4, 5, \ldots, \tag{5.8}$$

in the Borel plane z. In both φ^4 theory and QED, instanton and renormalon singularities lie on the negative and positive half-axes, respectively (see Fig. 7a); in QCD, the converse is true. The analysis presented above shows that factorial contributions due to particular diagrams arise in any field theory where the leading term in the expansion of β is quadratic.

It is obvious that 't Hooft's argumentation with regard to renormalons leaves unanswered the following basic questions: Why should certain sequences of diagrams be considered particularly important even though they comprise only a small fraction of all diagrams? How should we deal with double counting? (In other words, how do we know that renormalons are not taken into account in instanton contribution (2.5)?) However, the general question about the possibility of non-instanton contributions to the asymptotic behavior of expansion coefficients has been well reasoned, since it brings to light a shortcoming in the mathematical substantiation of Lipatov's method. Indeed, consider a function $f(x)$ that has a sharp peak at x_0 and a slowly decaying "tail" at large x (Fig. 6e), so that the contributions of the peak and tail regions to the integral $\int f(x)\,dx$ are comparable. An analysis of the integral would reveal the existence of a saddle point at x_0 and (if it is sufficiently sharp) show that the saddle-point method is formally applicable. However, the calculation of the integral in the saddle-point approximation would be incorrect, because the contribution of the tail

---

[10] Analogous singularities with $n = 0, -1, -2, \ldots$ (known as *infrared renormalons*) arise in the integral over the small-momentum region.

would be lost. If tails of this kind contribute to (2.4), then Lipatov's method fails.

Since there is hardly any alternative to the saddle-point method in calculations of functional integrals, direct analysis of possible tail contributions cannot be performed, and 't Hooft's argumentation is difficult to disprove. Nevertheless, it is "unnatural" in a certain sense: for any finite-dimensional integral (5.2), it can be shown that (a) its value for $g \longrightarrow 0$ is determined by saddle-point configurations [51] (according to (4.1.8), $g_c \longrightarrow 0$ as $N \longrightarrow \infty$) and (b) all singularities in the Borel plane are associated with action extrema ('t Hooft's argumentation based on (5.4) is necessary and sufficient). Therefore, renormalon singularities may arise only in the limit of an infinite-dimensional integral. However, the constructive argumentation in support of their existence is relatively weak and can easily be disproved by a careful analysis [48]. Further studies showed that summation of a more complicated sequences of diagrams leads to substantial modification of the renormalon contribution. The corresponding common coefficient becomes totally indeterminate [52]; i.e., the possibility that it vanishes cannot be ruled out. Thus, the existence of renormalon singularities is not an established fact, and this is admitted even by the most enthusiastic advocates of this hypothesis [53].

Nevertheless, 't Hooft's view immediately became popular [54–60]. This is explained by the use of diagrams in describing renormalons. One can easily insert a chain of bubbles into any diagram and explore the qualitative consequences of divergence of perturbation series for any phenomenon under study. Analysis of instantons can be combined with diagrammatic calculations (see [8]), but the procedure is very cumbersome. The concept of renormalons can undoubtedly be used as a "model." Moreover, conditions under which its use is justified can be determined (see Section 5.3). Regrettably, further investigation of high-order perturbative contributions was hampered after 't Hooft's lecture [50] had thrown doubt on the validity of Lipatov's method. As a consequence of the drop in its popularity, the complete perturbation-theory asymptotics in both QED and QCD remain uncalculated to this day even though all fundamental issues were resolved in the late 1970s.

5.2. Absence of Renormalon Singularities in φ^4 Theory

In the analysis of the Borel transforms arising in φ^4 theory presented in [48], it was shown that they are analytic on the complex plane with a cut extending from the nearest instanton singularity to infinity (see Fig. 7b), in agreement with a hypothesis (put forward by Le Guillou and Zinn-Justin [35]) that underlies an extremely efficient summation method, conformal–Borel technique (see Section 6.1). A comparison with 't Hooft's argumentation (see Fig. 7a) shows that all instanton singularities are absorbed by the cut, while renormalon singularities are absent.

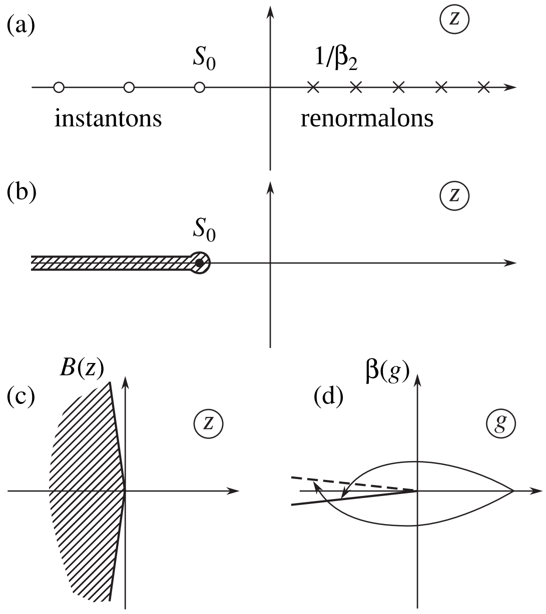


Fig. 7. (a) Singularities in φ^4 theory according to 't Hooft [50]. (b) Domain of analyticity according to [48]. Analyticity of $B(z)$ for $|\arg z| < \pi/2 + \delta$ (c) entails analyticity of $\beta(g)$ for $|\arg g| < \pi + \delta$, i.e., (d) on the entire physical sheet of the Riemann surface.

The approach developed in [48] relies on the use of the modified Borel transform in which $N!$ is replaced by $\Gamma(N + b_0)$ with an arbitrary b_0:

$$F(g) = \int_0^\infty dx e^{-x} x^{b_0 - 1} B(gx), \qquad B(g) = \sum_{N=0}^{\infty} \frac{F_N}{\Gamma(N + b_0)} g^N \tag{5.9}$$

(Borel–Leroy transform). It can readily be shown that all Borel transforms are analytic in the same domain, which is easy to find by setting $b_0 = 1/2$, since the corresponding Borel transform preserves exponential form:

$$F(g) = e^{-g} \longrightarrow B(z) = \frac{1}{2\sqrt{\pi}}\{e^{2i\sqrt{z}} + \text{c.c.}\}. \tag{5.10}$$

Accordingly, the Borel transform of (2.3) can be expressed as

$$B(z) = \frac{1}{2\sqrt{\pi}}\int D\varphi \exp(-S_0\{\varphi\}) \times [\exp(2i\sqrt{zS_{\text{int}}\{\varphi\}}) + \text{c.c.}], \tag{5.11}$$

and used to find the domain of analyticity for any finite-dimensional integral (2.3) calculated on a finite-size lattice for $m^2 > 0$.

The infinite-volume limit can be taken after Green functions are calculated as ratios of two integrals having the form of (2.3). This limit is not singular unless the system under study is at a phase-transition point, because partition into quasi-independent subsystems is possible when the correlation length is finite; in φ^4 theory, this possibility is guaranteed by the condition $m^2 > 0$.

Calculation of the continuum limit does not present any problem in the absence of ultraviolet divergences, which corresponds to $d < 2$. When the theory is divergent in the ultraviolet limit, the proof consists of the following steps:

(a) the domain of analyticity of $B(z)$ is determined for a finite cutoff parameter Λ by using Feynman regularization;

(b) the domain of analyticity is found for the Borel transform of the Gell-Mann–Low function and the anomalous dimensions defined in the cutoff scheme, whose dependence on Λ fades out as $\Lambda \longrightarrow \infty$;

(c) the invariance of the domain of analyticity under charge renormalization is proved;

(d) the domain of analyticity is determined for renormalized vertices and renormalization-group functions in other renormalization schemes.

Let us discuss a subtle detail of the proof that was not elucidated in [48]. Any quantity calculated perturbatively is a function of the bare charge g_B and the cutoff parameter Λ. Changing to a renormalized charge g, one obtains a function $F(g, \Lambda)$ that marginally depends on Λ, but approaches a finite limit as $\Lambda \longrightarrow \infty$ by virtue of renormalizability. Similarly, its Borel transform $B(z, \Lambda)$ tends to a finite limit $B(z)$. In [48], it was rigorously proved that $B(z, \Lambda)$ is analytic on the complex z plane with a cut extending from the nearest instanton singularity to infinity when Λ is finite. The function $B(z)$ is analytic in the same domain if the series is uniformly convergent (by the Weierstrass theorem [61]), which is the case when $B(z, \Lambda)$ is bounded (by compactness principle [62]). Therefore, regularity of $B(z)$ is guaranteed if the limit with respect to Λ is finite. However, renormalizability has been rigorously proved only in the framework of perturbation theory, i.e., for the coefficients of expansions in g and z, rather than directly for the functions $F(g, \Lambda)$ and $B(z, \Lambda)$. The proof presented in [48] assumes the existence of finite limits on the level of functions and is incomplete in this respect. However, the existence of these finite limits must be considered as a necessary physical condition for true renormalizability. This condition is directly related to the requirement of redefinition of functional integrals discussed below (see Section 5.3).

5.3. General Criterion for the Absence of Renormalon Singularities

The absence of renormalon singularities in the four-dimensional φ^4 theory, which is a typical “renormalon” theory, puts to question the general concept of renormalon. The problem of renormalons in an arbitrary field theory was elucidated in [63]. Returning to quantum electrodynamics, consider the simplest possible class of renormalon diagrams corresponding to all kinds of insertions into a photon line (see Figs. 6b and 6c). When the function β is known, all of these diagrams can easily be summed by solving the Gell-Mann–Low equation

$$\frac{dg}{d\ln k^2} = \beta(g) = \beta_2 g^2 + \beta_3 g^3 + \dots \tag{5.12}$$

under the initial condition $g(k^2) = g_0$ for $k^2 = m^2$ and analyzing the expansion of an integral having the form of (5.7) in terms of g_0.

The solution to Eq. (5.12) is

$$\begin{gathered} F(g) = F(g_0) + \ln\frac{k^2}{m^2}, \\ \text{where } F(g) = \int\frac{dg}{\beta(g)}. \end{gathered} \tag{5.13}$$

In view of the behavior of $F(g)$, the following expression can be used for small g:

$$\begin{gathered} F(g) = -\frac{1}{\beta_2 g} + f(g), \\ \text{if } \lim_{g \to 0} g f(g) = 0. \end{gathered} \tag{5.14}$$

The formal solution of (5.13) for g is

$$g(k^2) = F^{-1}\left\{-\frac{1}{\beta_2 g_0} + f(g_0) + \ln\frac{k^2}{m^2}\right\}. \tag{5.15}$$

Here, the right-hand side is regular at $g_0 = 0$; i.e., it can be represented as a series in powers of g_0 of the form

$$g = \sum_{N=1}^{\infty} A_N \left\{\frac{g_0}{r(x)}\right\}^N, \quad x = \beta_2 \ln\frac{k^2}{m^2}, \tag{5.16}$$

where $r(x)$ is the radius of convergence and A_N behaves as a power of N. The radius of convergence is determined by the distance to the singular point nearest to the origin.

If z_c is a singular point of the function $F^{-1}(z)$, then the singular points in g_0 in (5.15) satisfy the equation

$$z_c = -\frac{1}{\beta_2 g_0} + f(g_0) + \ln\frac{k^2}{m^2} \qquad (5.17)$$

or

$$g_0 x - 1 = \beta_2 g_0 [z_c - f(g_0)]. \qquad (5.18)$$

If z_c is finite and x is large, then Eq. (5.18) has a small root $g_0 \approx 1/x$. In this case, since the right-hand side is negligible by virtue of (5.14), there exists a singular point at $g_c \approx 1/x$, and the series in (5.16) is

$$\begin{aligned} g(k^2) &= \sum_{N=1}^{\infty} A_N (g_0 x)^N \\ &= \sum_{N=1}^{\infty} A_N \left(\beta_2 \ln\frac{k^2}{m^2}\right)^N g_0^N. \end{aligned} \qquad (5.19)$$

The integral obtained by substituting it into (5.7) is singular at the points defined by (5.8). (Note that the integral is dominated by the contributions of large k, which correspond to large x.) If $z_c = \infty$, then Eq. (5.17) has no solution for $g_0 \sim 1/x$, and the expansion coefficients in (5.19) must decrease faster than any exponential. Thus, the renormalon contribution is definitely smaller than the instanton contribution and the Borel plane does not contain any singularities.

If the function $z = F(g)$ is regular at g_0 and $F'(g_0) \neq 0$, then its inverse $g = F^{-1}(z)$ is also regular in some neighborhood of g_0. Therefore, the singular points of $F^{-1}(z)$ are $z_c = F(g_c)$, where g_c is any value such that either

$$F'(g_c) = 0 \ \text{ or } \ F'(g_c) \ \text{ does not exist.} \qquad (5.20)$$

In summary, renormalon singularities exist if there is at least one point g_c (including $g_c = \infty$) satisfying condition (5.20) and $z_c = F(g_c) < \infty$. Otherwise, renormalon singularities do not exist.

In terms of the β function, the results discussed above imply that renormalon singularities do not exist if $\beta(g) \sim g^\alpha$ with $\alpha \le 1$ at infinity and its singularities at finite g_c are so weak that the function $1/\beta(g)$ is nonintegrable at g_c (e.g., $\beta(g) \sim (g - g_c)^\gamma$ with $\gamma \ge 1$). When either condition is violated, there exist singular points defined by (5.8).

An analysis of more complicated classes of renormalon diagrams relying on the general Callan–Symanzik renormalization-group equation [63] leads to similar conclusions: necessary and sufficient conditions for the existence of renormalon singularities can be established, but no definite assertions can be made by using only results of renormalization-group analysis.

Now, recall that the perturbation series expansion of β is factorially divergent because there exists a cut emanating from the origin in the complex g plane. Therefore, both $g = 0$ and $g = \infty$ are branch points, and $\beta(g)$ can be represented by a Borel integral:

$$\beta(g) = \int_0^\infty dz e^{-z} B(gz) = g^{-1} \int_0^\infty dz e^{-z/g} B(z). \qquad (5.21)$$

Suppose that the Borel transform $B(z)$ behaves as z^α as z goes to infinity[11] (in which case $\beta(g) \propto g^\alpha$) and is a regular function for $|\arg z| < \pi/2 + \delta$, where $\delta > 0$ (see Fig. 7c). Then, $\beta(g)$ is a regular function for $|\arg g| < \pi + \delta$ (see Fig. 7d), which implies the absence of singularities at finite g in the physical sheet of the Riemann surface, and the behavior of β at infinity ($\beta(g) \sim g^\alpha$ with $\alpha \le 1$) guarantees the absence of renormalon singularities.

This criterion can be used as a basis for constructive analysis as follows. According to ’t Hooft (see Fig. 7a), instanton and renormalon singularities lie, respectively, on the negative and positive half-axes in both φ^4 theory and QED. If it is assumed that renormalon singularities do not exist, then (a) the regularity condition for $\beta(g)$ at finite g (see Figs. 7c and 7d) holds, (b) the asymptotic behavior of the expansion coefficients β_N is determined by the nearest instanton singularity and can be found by Lipatov’s method, and (c) the behavior of $\beta(g)$ at infinity can be uniquely determined by summing the corresponding perturbation series expansion since the Borel integral is well defined. If β grows faster than does g^α with $\alpha > 1$, then the starting assumption is incorrect, and the existence of renormalon singularities is proved by contradiction. If $\beta(g) \sim g^\alpha$ with $\alpha \le 1$, then the assumption that renormalons are negligible is self-consistent. These results are extended to QCD by changing the signs of g and z.

The program of determination of Gell-Mann–Low functions outlined here was implemented in [42, 47, 64, 65] (see discussion in Section 8). The exponent α is close to unity in both φ^4 theory and QED and much smaller than unity in QCD. Therefore, renormalon singularities can be self-consistently eliminated (up to uncertainty of results). Moreover, it can be argued that $\alpha = 1$ in both φ^4 theory and QED. Anyway, since $\beta(g)$ has a definite sign in both theories, the negligibility condition for renormalons is equivalent to the condition for their internal consistency.

---

[11] A similar analysis can be developed without using this assumption.

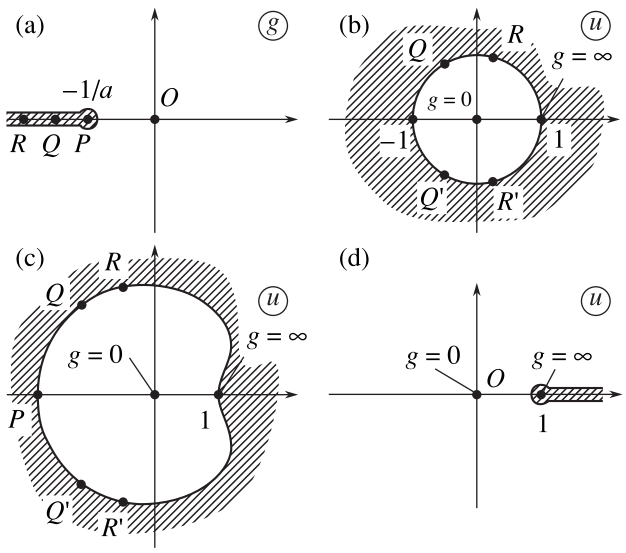


Fig. 8. Borel transform: (a) analyticity on the complex plane with the cut (–∞, –1/*a*); (b) analyticity domain conformally mapped to a unit circle. If analytic extension is restricted to the positive half-axis, then (c) conformal mapping can be performed to any domain such that $u = 1$ is its boundary point nearest to the origin and (d) to the plane with the cut (1, ∞) in an extreme case.

The only field theory in which the existence of renormalons is an established fact is the $O(n)$-symmetric sigma model in the limit of $n \longrightarrow \infty$ [53]. In this theory, the β function is calculated exactly in the one-loop approximation, and $\beta(g) \propto g^2$ at any g. Since $\alpha = 2$, renormalons cannot be eliminated self-consistently. However, the theory is self-contradictory in the four-dimensional case. It should be noted that renormalon singularities arise when a truncated series of any length is used to approximate the Gell-Mann–Low function; i.e., the "renormalon problem" cannot be solved in the framework of loop expansion [56, 57].

Note that functional integrals are ill defined if the existence of renormalon singularities cannot be ruled out. The conventional definition of functional integral based on perturbation theory is unsatisfactory because of the divergence of the expansion in terms of the coupling constant. Its adequate summation requires knowledge of certain analyticity properties in the Borel plane (see Section 6), which are indeterminate unless the existence of renormalon singularities is elucidated. The definition of functional integral as a multidimensional integral on a lattice also evokes doubts: a lattice theory can be qualitatively different from the continuum theory, because renormalon contributions correspond to arbitrarily large momenta. This leads to a deadlock: an analysis of functional integrals is required to solve the renormalon problem, but the integrals remain ill defined until the renormalon problem is solved. The proposed scheme of self-consistent elimination of renormalon singularities appears to be the only remedy. In this scheme, a continuum theory is interpreted as the limit of lattice theories by definition.[12]

6. PRACTICAL SUMMATION OF PERTURBATION SERIES

In this section, the practical summation of the following power series is discussed:

$$W(g) = \sum_{N=0}^{\infty} W_N (-g)^N, \tag{6.1}$$

where the expansion coefficients have the asymptotic form $ca^N\Gamma(N+b)$ and their numerical values are given. The present analysis is restricted to alternating series. Accordingly, $(-1)^N$ is factored out, and $a = -1/S_0 > 0$, as in φ^4 theory.

6.1. Conformal–Borel Technique and Other Methods

Treating (6.1) as Borel's sum (see Section 3), consider the following modification of Borel transform (5.9):

$$W(g) = \int_0^\infty dx e^{-x} x^{b_0 - 1} B(gx),$$
$$B(z) = \sum_{N=0}^{\infty} B_N (-z)^N, \quad B_N = \frac{W_N}{\Gamma(N+b_0)}, \tag{6.2}$$

where an arbitrary parameter b_0 can be used to optimize the summation procedure [35]. The Borel transform is assumed to have analyticity properties characteristic of φ^4 theory (see Section 5.2), i.e., analytic on the complex z plane with a cut extending from $-1/a$ to $-\infty$ (see Fig. 8a). The series expansion of $B(z)$ is convergent on the circle $|z| < 1/a$. To calculate the integral in (6.2), the series must be analytically continued. When the numerical values of W_N are given, continuation presents some difficulties. Its elegant solution proposed in [35] makes use of the conformal mapping $z = f(u)$ of the plane with

---

[12] This philosophy is hidden in the very concept of renormalizability. In fact, an effective theory is constructed for small momenta with a cutoff parameter Λ, and a similar scale is supposed to be the upper boundary of the essential domain of integration. Renormalizability is interpreted as a possibility of taking the limit as $\Lambda \longrightarrow \infty$ without corrupting the structure of the theory. However, the actual contribution of momenta larger than Λ can hardly be controlled within the scope of the effective theory, since an essential contribution may always come from some "demon" that lurks in the large-momentum region, running away with increasing Λ. The contribution of the action minimum associated with a lattice instanton in the theory of the Anderson transition is a remarkable example of physical realization of such a demon [8, 36, 37].

a cut to the unit circle $|u| < 1$ (see Fig. 8b) to represent $B(z)$ as a series in powers u,

$$B(z) = \sum_{N=0}^{\infty} B_N(-z)^N\Big|_{z=f(u)} \longrightarrow B(u) = \sum_{N=0}^{\infty} U_N u^N, \tag{6.3}$$

which is convergent at any z. Indeed, the singular points of $B(z)$ (P, Q, R, …) lie on the cut, while their images (P, Q, Q', R, R', …) lie on the circumference of the circle $|u| = 1$. Thus, the latter series in (6.3) is convergent at $|u| < 1$. However, there is one-to-one correspondence between the interior $|u| < 1$ of the circle and the domain of analyticity in the z plane. The conformal mapping is defined as

$$z = \frac{4}{a}\frac{u}{(1-u)^2} \quad \text{or} \quad u = \frac{(1+az)^{1/2}-1}{(1+az)^{1/2}+1}. \tag{6.4}$$

Hence,

$$U_0 = B_0, \qquad U_N = \sum_{K=1}^{N} B_K\left(-\frac{4}{a}\right)^K C_{N+K-1}^{N-K} \quad (N \geq 1), \tag{6.5}$$

where C_N^K denotes binomial coefficients.

Since the $B(z)$ exhibits power-like behavior at infinity, the Borel integral in (6.2) is rapidly convergent, and its upper limit can be assigned a finite value in accordance with the required accuracy. Then, u is bounded from above by $u_{\max} < 1$, and the latter series in (6.3) is convergent. The substitution of $u = u(z)$ into the Borel integral (6.2) followed by integration, as done in [35], is not quite legitimate, because permutation of the sum and the integral may lead to slower convergence or even divergence of the algorithm. However, this scheme is convergent, because the actual coefficients U_N exhibit power-like asymptotic behavior (see [65, Section 2.1]).[13] As b_0 increases, oscillatory asymptotic behavior changes to monotonic, and this change was used in [35] as a basis for error estimation.

When the first N_m coefficients in series (6.1) are known, formula (6.5) can be used to find the first N_m coefficients of the convergent series in (6.3). If $g \sim 1$, then the dominant contribution corresponds to values of u on the order of a few tenths, which makes it possible to obtain accurate results even for small N_m. The summation error is estimated as

$$\delta W(g) \propto \exp\{-3(N_m^2/ag)^{1/3}\}, \quad ag \lesssim N_m^2. \tag{6.6}$$

Therefore, the highest value of g for which satisfactory results can be obtained is on the order of N_m^2. This scheme was used [35] to calculate critical exponents in theory of phase transitions up to the third decimal place.

In an alternative analytic continuation method [66], the coefficients in the Padé approximant $[M/L]$ (defined as the ratio $P_M(z)/Q_L(z)$ of polynomials of degrees M and L) are adjusted to approximate the function $B(z)$ so that the known lowest order terms in (6.1) are obtained. Diagonal and quasi-diagonal approximants (with $M = L$ and $M \approx L$, respectively) are known to converge to the approximated function as $M \longrightarrow \infty$ for a broad class of functions, but the convergence rate is relatively low in the general case. When both M and L are finite, the approximate Borel transform generally exhibits incorrect behavior at infinity dictated by the particular Padé approximant $[M/L]$ employed. The corresponding incorrect behavior of $W(g)$ at $g \longrightarrow \infty$ gives rise to a certain error at $g \sim 1$ by continuity. Accordingly, different results are obtained by using different approximants, depending on the subjective choice of the user. However, uniform convergence with respect to g can be achieved by matching the Padé approximant at infinity with the asymptotic form of $B(z)$ if the asymptotic behavior of $W(g)$ in the strong-coupling limit is known. When the number of terms in the expansion is sufficiently large, the asymptotic behavior at strong coupling can be “probed” by analyzing the convergence rate, as done in [66]. Furthermore, information concerning the singular behavior at $-1/a$ can be used to restrict the approximation by requiring that all poles of Padé approximants lie on the negative half-axis, in which case the input data used in this method are the same as in the conformal–Borel technique described above. The results of the original calculations of critical exponents performed by this method in [66] were virtually identical with those obtained in [35]. The Padé–Borel technique should be used when the analyticity properties of $B(z)$ are not known, because information about the locations of the nearest singularities can be gained by constructing Padé approximants.

In multiple-charge models, Chisholm approximants (rational functions in many variables) can be used instead of Padé approximants [67]. A more efficient approach to problems of this type is based on the so-called resolvent expansion [68]: all charges are multiplied by an auxiliary parameter λ, Padé approximants in terms of this parameter are employed, and λ is set to unity at the end of the calculations. In this method, the symmetry of the model is completely preserved, and projection onto any charge subspace of lower dimen-

---

[13] To determine the asymptotic form of U_N, the contributions found in [65] must be summed over all singular points, whose number is infinite. Since the sum is finite for any constant $N = N_0$, it can readily be shown that it is dominated by the term containing the highest power of N as $N \longrightarrow \infty$.

sion does not lead to loss of information [69, 70]. A more complicated sequence of approximants can be constructed by using Winn's ϵ-algorithm [71] based on a "strong" Borel transform (see Section 7). The Sommerfeld–Watson summation scheme [13] makes use of the analyticity properties of the coefficient functions of Borel transforms. A generalized conformal–Borel technique was employed in [12]. The last two methods make it possible to "guess" the strong-coupling asymptotics analyzed systematically in the next section.

In another approach, variational perturbation theory [72, 73] is used to formulate a scheme of interpolation between the weak- and strong-coupling regions when some information about the latter is available. With regard to critical exponents, this information concerns behavior near the renormalization-group fixed point. It can be expressed in terms of strong coupling as an expansion in the bare charge. Thus, a divergent perturbation series is transformed into a convergent sequence of approximations, and an accuracy comparable to that of conformal–Borel technique is achieved [74]. However, neither divergence of the series nor the Lipatov asymptotic form is used in this approach explicitly. Information about the latter can be used only implicitly by interpolating the coefficient function. Since an attempt to this effect made in [75] did not result in any improvement in accuracy, Kleinert claimed that information about high-order terms is insignificant. It is obvious that this assertion is incorrect in the general case: since exact knowledge of the expansion coefficients is equivalent to exact knowledge of the function, appropriate use of any additional information must improve accuracy. The particular result obtained by Kleinert is due to the fact that variational interpolation is not less accurate than interpolation of the coefficient function.

The application of divergent series to calculation of critical exponents relies on the use of the Callan–Symanzik renormalization-group equation, which contains both Gell-Mann–Low function $\beta(g)$ and renormalization-group functions $\eta(g)$ and $\eta_4(g)$ (anomalous dimensions) [36, 76, 77]. If g^* is a nontrivial zero of $\beta(g)$, then the critical exponents η and ν can be expressed in terms of the anomalous dimensions at that point by the relations $\eta(g^*) = \eta$ and $\eta_4(g^*) = 1/\nu - 2 + \eta$, and the remaining exponents are determined by well-known relations [39]. The renormalization-group functions are calculated as series expansions in terms of g, which can be summed by methods mentioned above.

Based on this approach, substantial progress has been made in analysis of critical behavior of a various systems. The critical exponents for the $O(n)$-symmetric φ^4 theories with $n = 0$, 1, 2, and 3 were originally calculated in [35, 66] for two- and three-dimensional spaces by using six- and four-loop expansions, respectively. Subsequently, these calculations were extended to larger n and higher order coupling constants [78, 79]. In particular, this provided a basis for estimating the computational scope of the $1/n$ expansion. The seven-loop contributions to the renormalization-group functions found for $d = 3$ in [75, 80] and for $d = 2$ in [81, 82] were used to refine the critical exponents. The latter studies revealed systematic deviation of the resummation results from the known exact values of critical exponents. However, their interpretation as a manifestation of nonanalytic contributions to renormalization-group functions (see [81, 82]) does not seem to be well grounded.

When cubic anisotropy is taken into account, a two-charge version of the n-component φ^4 theory is obtained, since the corresponding action contains two fourth-order invariants. The cases of $n = 3$ and $n = 0$ correspond to a cubic crystals and weakly disordered Ising ferromagnets, respectively. This can be shown by using the standard replica trick to average over the random impurity field [83]. The expansion coefficients and sums of renormalization-group series were calculated for these systems in the four-, five-, and six-loop approximations in [69, 84], [85], and [86, 87], respectively (see also [88, 89]).

Six-loop expansions have also been obtained for other two-charge field systems: the mn model describing certain magnetic and structural phase transitions (including the critical behavior of n-component disordered magnets in the case of $m = 0$ [87]) and the $O(m) \times O(n)$-symmetric model corresponding to the so-called chiral phase transitions [90]. The summation of the resulting series performed in [90–93] made it possible to elucidate the structure of the phase portraits of the renormalization-group equations and analyze the stability of nontrivial fixed points. Even more complicated (three-charge) versions of φ^4 theory arise in models of superconductors with nontrivial pairing, many-sublattice antiferromagnets, structural phase transitions, superfluid transition in neutron liquid, etc. Some of these have been analyzed in three-, four-, and six-loop approximations in [70, 94–96]. In recent studies, five-loop expansions were found and resummed for the two-dimensional chiral [97], cubic, and mn models [82]. Finally, summation of three- and four-loop expansions was used to analyze critical dynamics in pure and disordered Ising models [98, 99], as well as effects due to long-range interactions [100] and violation of replica symmetry [101].

Note that, instead of summing a series expansion in terms of the coupling constant in the space of physical dimension, one can sum up the divergent ϵ-expansions obtained for the formal problem of phase transition in the $(4 - \epsilon)$-dimensional space [77, 102, 103]. The four-loop expansions for the $O(n)$-symmetric theory [104] summed up in [105] were extended to the five-loop level [106] and summed in [80]. The five-loop expansions obtained in [107] for the cubic model were used in [108] as a basis for deriving five-loop $\sqrt{\epsilon}$-expansions of critical exponents for disordered Ising model, and their summation was discussed in [108–110].

Detailed discussions of the current status of theory of critical phenomena based on multiple-loop renormalization-group expansions, as well as extensive bibliography, can be found in recent reviews [89, 96, 111].

6.2. Summation in the Strong-Coupling Limit

The results obtained for increasingly stronger coupling are characterized by stronger dependence on the particular implementation of the summation procedure. Analysis of the ensuing uncertainty is facilitated if the asymptotic behavior of $W(g)$ in the strong-coupling limit are directly related to the values of W_N. This problem is solved here by assuming that the asymptotic behavior can be represented by the power law

$$W(g) = W_\infty g^\alpha, \quad g \longrightarrow \infty, \tag{6.7}$$

which adequately represents all models that are amenable to analysis and consistent with realistic field-theoretic problems. The problem can be solved by applying the standard conformal–Borel technique (see Section 6.1) [65]. However, a more efficient algorithm can be developed by using a modified conformal mapping.

If $z = 0$ and ∞ are mapped to $u = 0$ and 1, respectively, then (6.4) is the only conformal mapping that can be used to find the analytic continuation of the Borel transform to arbitrary complex z. However, this is not necessary: the integral in (6.2) can be calculated if the analytic continuation to the positive half-axis is found. Therefore, it will suffice to use a conformal mapping to any region for which $u = 1$ is the boundary point nearest to the origin (see Fig. 8c): under this condition, the latter series in (6.3) will be divergent if $|u| < 1$, in particular, on the interval $0 < u < 1$ (the image of the positive half-axis).

One advantage of this conformal mapping is that the divergence of the re-expanded series in (6.3) is controlled by the nearest singular point $u = 1$, which is related to the singularity of $W(g)$ as $g \longrightarrow \infty$, so that the asymptotic form of U_N corresponds to the parameters of asymptotic formula (6.7). If U_N is expressed in terms of $B(u)$ as

$$U_N = \oint_C \frac{du}{2\pi i}\frac{B(u)}{u^{N+1}}, \tag{6.8}$$

and the contour C encompassing $u = 0$ is deformed so that it goes around the cuts extending from the singular points to infinity, then it can readily be shown that the asymptotic form of U_N for large N is controlled by the nearest singular point $u = 1$. To reduce the contributions of the remaining singular points $P, Q, Q', \ldots$, the points should be moved away as far as possible. Ultimately, the desired transform is the mapping to the plane with the cut $(1, \infty)$ (see Fig. 8d),

$$z = \frac{u}{a(1-u)}, \tag{6.9}$$

for which

$$U_0 = B_0,$$
$$U_N = \sum_{K=1}^{N} \frac{B_K}{a^K}(-1)^K C_{N-1}^{K-1} \quad (N \geq 1). \tag{6.10}$$

The asymptotic form of U_N for large N,

$$U_N = U_\infty N^{\alpha-1}, \quad N \longrightarrow \infty, \tag{6.11}$$

$$U_\infty = \frac{W_\infty}{a^\alpha \Gamma(\alpha)\Gamma(b_0+\alpha)}, \tag{6.12}$$

is determined by the parameters of asymptotic formula (6.7). Thus, a simple algorithm is proposed: use (6.2) to calculate the coefficients B_N corresponding to the prescribed W_N; substitute the results into (6.10) to find U_N; and find power-law fit (6.11) to determine the parameters W_∞ and α in (6.7).

In practical applications of the algorithm, one has to deal with problems due to random error growth. The random error in U_N corresponding to a relative computational or round-off error δ in W_N is an extremely rapidly increasing function of N:

$$\delta U_N \sim \delta \cdot 2^N. \tag{6.13}$$

In double-precision computations, when $\delta \sim 10^{-14}$, the value of δU_N is comparable to unity if $N \approx 45$, and the corresponding error in the recovered asymptotic formula (6.7) is ~1%.

Fortunately, the algorithm is "superstable" with respect to smooth errors in the sense that the output error is smaller than the input error. Linear mapping (6.10) has a remarkable property:

$$\sum_{K=1}^{N} K^m (-1)^K C_{N-1}^{K-1} = 0 \tag{6.14}$$

for $m = 0, 1, \ldots, N-2$. Accordingly, the addition of an arbitrary polynomial $P_m(K)$ to B_K/a^K (which behaves as a power of K) does not change the asymptotic form of U_N. There is an analogous property characteristic of a broad class of smooth functions accurately approximated by polynomial fitting: in particular, when B_K/a^K is replaced with $B_K/a^K + f(K)$, where $f(K)$ is an entire function with rapidly decreasing Taylor series expansion coefficients, the resulting change in U_N rapidly decreases with increasing N. In practical problems, sev-

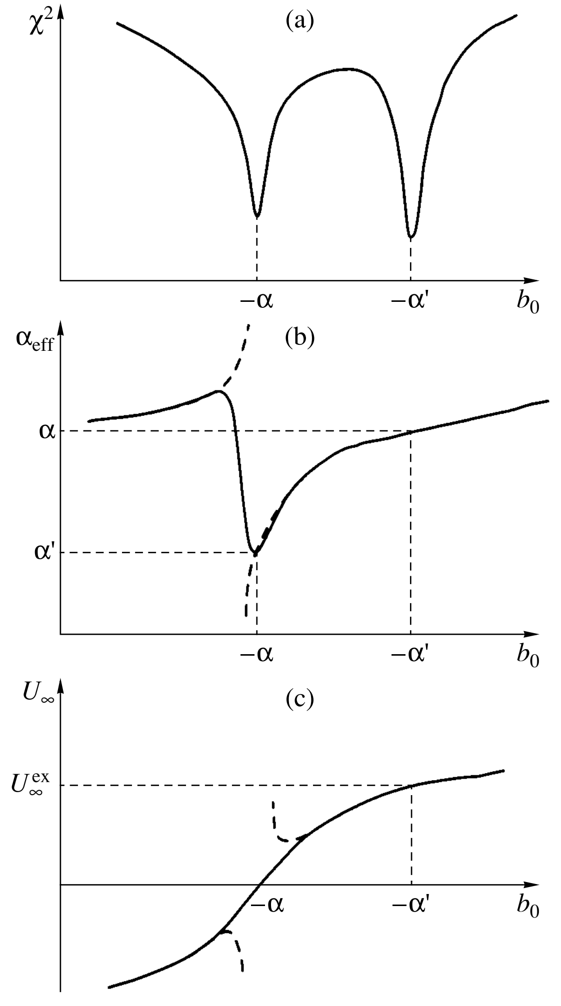


Fig. 9. Theoretical curves of χ^2, $\alpha_{\rm eff}$, and U_∞ plotted versus b_0 by neglecting the corrections subsumed under the ellipsis in (6.16). Log–log plots are frequently discontinuous (dashed curves) [65].

eral low-order values and asymptotic behavior of W_N are known, while intermediate coefficients are found by interpolation. Since interpolation leads to smooth errors, they can be expected to play a minor role even if their magnitudes are large. Thus, the proposed algorithm is best suited to the most relevant formulation of the problem.

The approximation of U_N by a power law can be performed by applying a standard χ^2 minimization procedure [112], and the fitting interval $N_{\min} \le N \le N_{\max}$ can be chosen as follows. The upper limit $N_{\max}$ is set by imposing the condition $\delta U_N \sim U_N$, because no additional information can be derived from coefficients with higher N. This condition is not restrictive, because outliers are automatically discarded in χ^2 minimization. The lower limit $N_{\min}$ is set by requiring that χ^2 have "normal" values to rule out the systematic error due to deviation of U_N from asymptotic law (6.11).

The existence of an upper limit for N entails strong dependence of the results on b_0 (see (6.2)), because this parameter determines the rate at which the asymptotic behavior is approached. To analyze this effect, suppose that asymptotic formula (6.7) is modified by adding power-law corrections:

$$W(g) = W_\infty g^\alpha + W'_\infty g^{\alpha'} + W''_\infty g^{\alpha''} + \dots \qquad (6.15)$$

By analogy with (6.11) and (6.12), it follows that

$$U_N = \frac{W_\infty}{a^\alpha \Gamma(\alpha)\Gamma(b_0+\alpha)} N^{\alpha-1} + \frac{W'_\infty}{a^{\alpha'}\Gamma(\alpha')\Gamma(b_0+\alpha')} N^{\alpha'-1} + \dots \qquad (6.16)$$

When the corrections subsumed under the ellipsis in (6.16) are neglected, the formal approximation of (6.16) by power law (6.11) leads to satisfactory results, because the log–log plot of (6.16) is almost linear; however, the values of α and U_∞ thus obtained should be interpreted as "effective" parameters.

Since the first and second terms in (6.16) vanish at the poles of the respective gamma functions, $U_N \propto N^{\alpha'-1}$ and $U_N \propto N^{\alpha-1}$ are obtained for $b_0 = -\alpha$ and $b_0 = -\alpha'$, respectively. The corresponding power-law fits are particularly accurate (the values of χ^2 are low). The results obtained by varying b_0 are illustrated by Fig. 9. The graph of χ^2 has two sharp minima at $b_0 = -\alpha$ and $b_0 = -\alpha'$. The curve of $\alpha_{\rm eff}$ drops to α' in the neighborhood of $b_0 = -\alpha$ and approaches α outside this neighborhood. At $b_0 = -\alpha'$, the exact equality $\alpha_{\rm eff} = \alpha$ is reached and effective parameter U_∞ exactly corresponds to W_∞. In the neighborhood of $b_0 = -\alpha$, U_∞ vanishes, while its linear slope

$$U_\infty \approx \frac{W_\infty}{a^\alpha \Gamma(\alpha)}(b_0+\alpha) \qquad (6.17)$$

yields an estimate for W_∞ weakly sensitive to errors in α. The effect of the terms discarded in (6.16) only slightly changes this pattern.

The analysis above suggests that independent estimates for the exponent α can be obtained by using (i) the value of $\alpha_{\rm eff}$ at the right-hand minimum of χ^2, (ii) the location of the left-hand minimum of χ^2, (iii) a log–log plot in the neighborhood of zero U_∞, and (iv) power-law fitting in this neighborhood with a constant α given by a previous estimate.

Similarly, independent estimates for W_∞ are obtained by using the value of U_∞ at the right-hand minimum of χ^2 and the linear slope of $U_\infty(b_0)$ near its zero. In the latter estimate, the constant α is varied within the interval of its values estimated as described above to obtain upper and lower bounds for W_∞.

The accuracy of the results can be estimated by using the fact that the discrepancies between different estimates for α and W_∞ are comparable to their respective deviations from the exact value. When several estimates are available, the results are more reliable: whereas two close estimates may be obtained accidentally (resulting in underestimation of the error), proximity of three or four independent estimates is very unlikely.

As an example, consider the integral

$$W(g) = \int_0^\infty d\varphi \exp(-\varphi^2 - g\varphi^4), \tag{6.18}$$

which can be interpreted as the zero-dimensional analog of the functional integral in φ^4 theory. Its asymptotic behavior is described by (6.7) with $\alpha = -1/4$ and $W_\infty = \Gamma(1/4)/4$, and the corrections can be represented as a power series in $g^{-1/2}$. The results obtained when W_N are double-precision quantities ($\delta \sim 10^{-14}$) are illustrated by Fig. 10.[14] The ensuing estimates,

$$\begin{gathered} \alpha = -0.235 \pm 0.025, \quad W_\infty = 0.908 \pm 0.025, \\ \alpha' = -0.75 \pm 0.08, \end{gathered} \tag{6.19}$$

are in good agreement with the respective exact values

$$\begin{gathered} \alpha = -0.25, \quad W_\infty = 0.9064\ldots, \\ \alpha' = -0.75. \end{gathered} \tag{6.20}$$

Stability of the algorithm with respect to interpolation can be checked by writing the expansion coefficients as[15]

$$W_N = ca^N \Gamma(N+b) \times \left\{1 + \frac{A_1}{N} + \frac{A_2}{N^2} + \ldots + \frac{A_K}{N^K} + \ldots\right\}, \tag{6.21}$$

since relative corrections to the Lipatov asymptotic form have the form of a regular series expansion in terms of $1/N$. This representation can readily be used to interpolate the coefficient function: the series can be truncated, and the parameters A_K can be determined by comparison with several low-order coefficients W_N. When interpolation is performed by using the Lipatov asymptotic parameters, the lowest order correction A_1/N, and the single coefficient W_1 [65], the resulting values $\alpha = -0.245 \pm 0.027$ and $W_\infty = 0.899 \pm 0.014$ are almost equal to those in (6.20). The errors in these results are still determined by round-off error, even though the interpolation errors are greater by ten orders of magnitude.

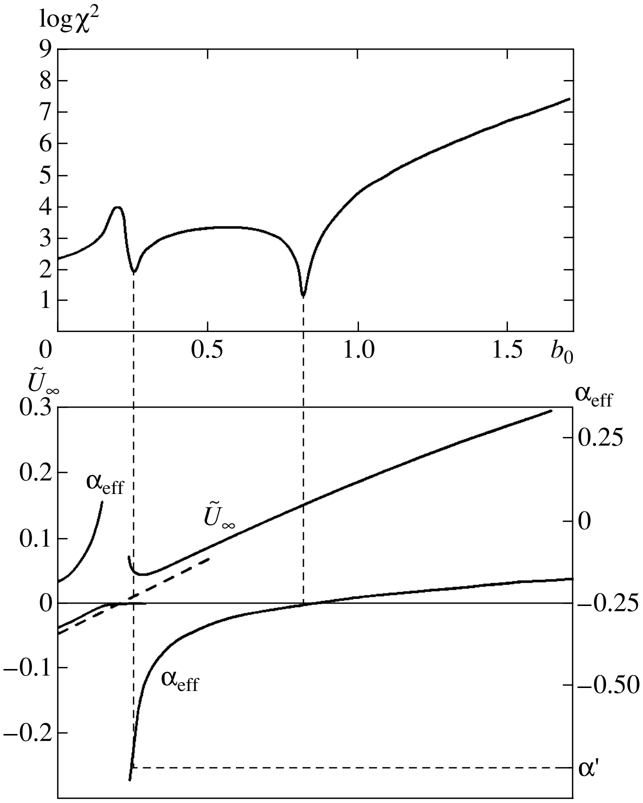


Fig. 10. Curves of χ^2, $\alpha_{\rm eff}$, and $\tilde{U}_\infty$ calculated as functions of b_0 for integral (6.18) by using the interval $24 \le N \le 50$. Dashed curve represents $\tilde{U}_\infty(b_0)$ in the neighborhood of its zero by using $\alpha = -0.25$.

Another example is the calculation of the ground-state energy $E_0(g)$ for anharmonic oscillator (1.1). This model can be reformulated as φ^4 theory. The parameters of the asymptotic power law obtained in [65] by using the coefficients W_N calculated in [6] up to $\delta \sim 10^{-12}$,

$$\alpha = 0.317 \pm 0.032, \quad W_\infty = 0.74 \pm 0.14, \tag{6.22}$$

agree with the exact values $\alpha = 0.3333\ldots$ and $W_\infty = 0.6679\ldots$ and demonstrate adequately estimated errors.

The reliability of error estimation suggests a new approach to optimization of summation algorithms. On a conceptual level, optimization is performed by introducing a variation characterized by a parameter λ into the summation algorithm and finding the parameter value that maximizes the convergence rate. For exam-

---

[14] For technical reasons, the coefficients $\tilde{U}_N = U_N \Gamma(b_0 + N_0)$ given below are normalized so that a constant limit is obtained as $b_0 \longrightarrow \infty$. Here, N_0 is the lower limit of summation in (6.5), which may differ from unity if the first terms in (6.1) vanish. Similarly, $\tilde{U}_\infty = U_\infty \Gamma(b_0 + N_0)$.

[15] Frequently arising questions concerning the analyticity of the coefficient function and its interpolation were discussed in [113] in the context of comments to [114].

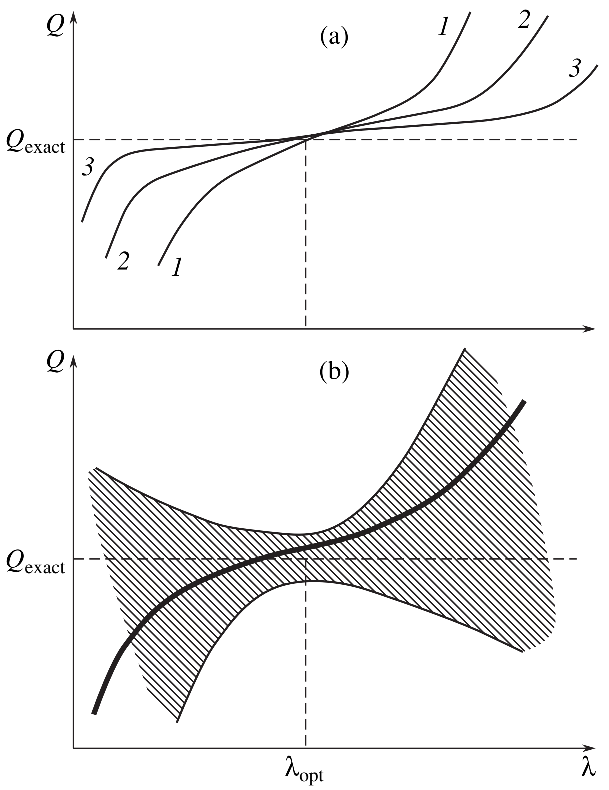


Fig. 11. (a) In theory, any quantity Q obtained by summation of a series is independent of the optimization parameter λ. In practice, the dependence exists and changes from curve *1* to curves *2* and *3* with increasing amount of available information. The optimal value of λ corresponds to the central plateau region. (b) The error of approximate calculation of Q (hatched region) depends on λ. When the error is estimated correctly, the exact value Q_{exact} is consistent with all data. In the "ideal" case illustrated here, the optimal value of λ corresponds to the minimal error.

ple, series (6.1) can be raised to the power λ, and the summation algorithm can then be applied to the reexpanded series

$$W(g)^{\lambda} = \tilde{W}_0 - \tilde{W}_1 g + \tilde{W}_2 g^2 - \dots + \tilde{c} a^N \Gamma(N+b)(-g)^N + \dots, \tag{6.23}$$

which has analogous properties (except for a different parameter c in the Lipatov asymptotic form [8]). Theoretically, series (6.23) is equivalent to (6.1) in that the value of any quantity Q obtained as a result of summation must be independent of λ. When the available information about series (6.1) is incomplete, Q depends on λ. The dependence becomes weaker as the amount of information increases. In the general case, convergence is not uniform with respect to λ, and the approximate value of Q is close to the exact one only within some plateau-like region (see Fig. 11a). The error rapidly increases away from it, while the plateau widens and flattens with increase in available information (e.g., see [115]). It is clear that the best convergence is achieved at the center of the plateau. However, the location of a "center" may not be easy to determine, since the plateau may be asymmetric or indistinct, its center may move in the course of interpolation, etc. Therefore, the choice of λ, i.e., the optimization of Q, is a largely subjective one.

However, there exists an objective approach to optimization. Note that the λ determines not only the approximate value of Q, but also the error of its calculation. If the error is estimated correctly, then the exact value Q_{exact} must be consistent with the approximate results corresponding to any λ (see Fig. 11b); i.e., spurious dependence of Q on λ is ruled out. If this "ideal" is attained, then optimization of λ reduces to error minimization.

It is reasonable to optimize the interpolation stage, because the uncertainty of results is ultimately due to imprecise knowledge of the coefficients W_N. If (6.21) is rewritten as

$$W_N = c a^N N^{\tilde{b}} \Gamma(N + b - \tilde{b}) \left\{ 1 + \frac{\tilde{A}_1}{N - \tilde{N}} + \frac{\tilde{A}_2}{(N - \tilde{N})^2} + \dots + \frac{\tilde{A}_K}{(N - \tilde{N})^K} + \dots \right\} \tag{6.24}$$

and the interpolation is performed by truncating the series and determining the coefficients $\tilde{A}_K$, then the interpolation procedure can be parameterized by $\tilde{b}$ and $\tilde{N}$. Optimization with respect to $\tilde{b}$ relies on a theoretical analysis [65], and the result $\tilde{b} = b - 1/2$ corresponds to the Lipatov asymptotic form parameterized as $ca^N N^{b-1/2}\Gamma(N + 1/2)$. Optimization with respect to $\tilde{N}$ was demonstrated in [65], where interpolation over the first nine coefficients was performed for the series corresponding to anharmonic oscillator. Coarse optimization of χ^2 as a function of $\tilde{N}$ was performed for several constant values of b_0 having minima at $\tilde{N}$ between –5.5 and – 5.0. This narrow interval determines the range of interpolations consistent with the power-law asymptotic behavior of $W(g)$. Next, a systematic procedure was executed to find α and W_∞. The "ideal" situation illustrated by Fig. 11b was obtained by widening the error corridor for α by a factor of 1.3 and for W_∞ by a factor of 1.1, which can be done since the error is estimated up to order of magnitude. If the values of α and W_∞ are consistent with all data, and the one-sided error is minimized, the results are

$$\alpha = 0.38 \pm 0.05, \quad W_\infty = 0.52 \pm 0.12, \tag{6.25}$$

and their deviation from the exact results is adequately estimated by the respective errors.

If the available information concerning W_N is sufficient to recover the asymptotic behavior of $W(g)$, then the summation of series (6.1) presents no problem: the lowest order coefficients U_N are calculated by using (6.10) and continued to match the asymptotic form $U_\infty N^{\alpha-1}$. As a result, all coefficients in the convergent series (6.3) are found. The summation error is determined by the accuracy of calculation of the asymptotic form of U_N, which is characterized by a quantity Δ assumed to be constant within a bounded interval even though the actual error behaves as a logarithm of N. If N_c is the characteristic value for which the relative error is equal to Δ and

$$\frac{\delta U_N}{U_N} = \begin{cases} 0, & N < N_c, \\ \Delta, & N \ge N_c, \end{cases} \tag{6.26}$$

then the summation error is

$$\frac{\delta W(g)}{W(g)} \sim \begin{cases} \Delta, \quad ag \gtrsim N_c, \\ \Delta \exp\{-2(N_c/ag)^{1/2}\}, \quad ag \lesssim N_c. \end{cases} \tag{6.27}$$

The error of a straightforward summation using N_m known coefficients is given by (6.27) with $N_c = N_m$ and $\Delta \sim 1$ and is higher than estimated by (6.6). Nevertheless, the stability of the algorithm with respect to interpolation ensures that $\Delta \ll 1$ and $N_c \gg N_m$ even for small N_m [65].

7. "NON-BOREL-SUMMABLE" SERIES

It is clear from Section 3 that any definition of the sum of a factorially divergent series must be equivalent to Borel's definition. In other words, any effective summation scheme must be equivalent to the Borel technique; otherwise, self-consistent manipulation of divergent series is impossible. Nevertheless, "non-Borel-summable" series are frequently discussed in the literature. This misleading concept is used in two situations.

In one of these situations, the coefficients of the series in question increase much faster than $N!$, and the standard transform defined by (3.11) is not effective, but the "strong" Borel transform

$$\begin{aligned} F(g) &= \sum_{N=0}^{\infty} F_N g^N \\ &= \sum_{N=0}^{\infty} \frac{F_N}{\Gamma(kN+1)} \int_0^{\infty} dx x^{kN} e^{-x} g^N \\ &= \int_0^{\infty} dx e^{-x} \sum_{N=0}^{\infty} \frac{F_N}{\Gamma(kN+1)} (gx^k)^N, \end{aligned} \tag{7.1}$$

can be used for summing series whose coefficients increase as $(N!)^k$ with arbitrary finite k.[16]

In the other situation, the factorial series in question have nonalternating coefficients. Analysis of the simple example

$$F(g) = \sum_{N=0}^{\infty} a^N N! g^N = \int_0^{\infty} dx \frac{e^{-x}}{1-agx}, \quad ag > 0, \tag{7.2}$$

shows that the singular points of the corresponding $B(z)$ lie on the positive half-axis; i.e., they belong to the contour of integration in (3.12). Therefore, the Borel integral is ill defined and must be correctly interpreted. In particular, the contour of integration in (7.2) may lie above or below the singular point $x = 1/ag$; alternatively, the integral may be interpreted in the sense of the Cauchy principal value.

To deal with all possible interpretations, the definition of gamma function is written as

$$\Gamma(z) = \sum_i \gamma_i \int_{C_i} dx e^{-x} x^{z-1}, \quad \sum_i \gamma_i = 1, \tag{7.3}$$

where $C_1, C_2, \ldots$ are arbitrary contours extending from the origin to infinity in the right half-plane. Then, the Borel transform becomes

$$F(g) = \sum_i \gamma_i \int_{C_i} dx e^{-x} x^{b_0-1} B(gx), \tag{7.4}$$

where the contours C_i are not mutually equivalent because of the singularities of the Borel transform $B(z)$ and cannot be aligned with the positive half-axis as can be done in (7.3). The choice of interpretation is determined by the parameters γ_i if the set $\{C_i\}$ contains all nonequivalent contours.

Correct interpretation of the Borel integral is impossible without additional mathematical information about the quantity represented by a divergent series. For this reason, current views on the prospects of recovering physical quantities from the corresponding perturbative

---

[16] This leads one to the following question: why do we not use "strong" transforms of this kind in every case whatsoever? As far as exact calculations are concerned, the only criterion is analytical tractability: any transform is applicable if the required calculations can be carried through. In approximate calculations, "strong" summation methods are not as reliable as they may seem to be. Mathematically, the nontriviality of a function is determined by the type and location of its singularities. Strong Borel transform (7.1) defines an entire function $B(z)$ having a complicated singular point at infinity. This very fact entails practical difficulties: the singular point is hardly amenable to analysis, whereas its impact is not any weaker. This explains why the "weakest" Borel transform is so effective: its singularities lie at finite points in the complex plane, and even their location provides essential information about the function (see Section 6).

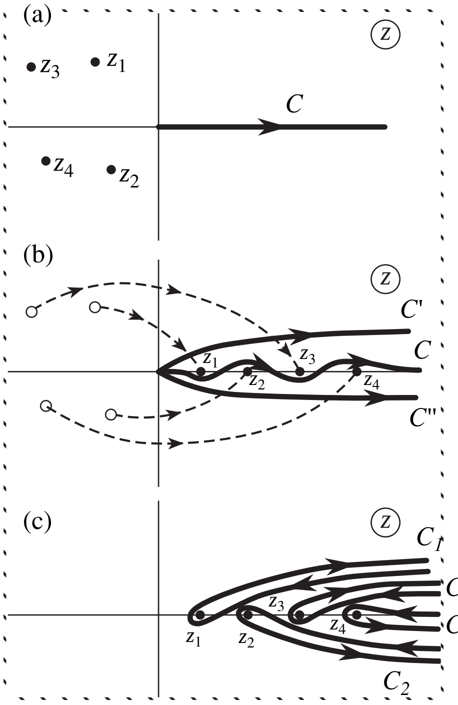


Fig. 12. Graphic illustration of non-Borel-summability.

expansions are largely pessimistic (see [116, 117]). However, the importance of additional information should not be overestimated. Reasonable interpretation can be based on the requirement of analyticity with respect to model parameters, which is satisfied in almost every physical application.

If the values of the parameters are such that all singular points z_i of the Borel transform lie in the left half-plane (see Fig. 12a), then Borel integral (3.12) for a factorial series with asymptotic expansion coefficients $ca^N\Gamma(N + b)$ is the only analytic function that satisfies the strong asymptotic condition within the sector $|\arg g| \le \pi$, $|g| < g_0$ with an arbitrary g_0 (see [21, Section 8.1]), i.e., on the entire physical sheet of the Riemann surface. Therefore, the choice of a contour C aligned with the positive half-axis (see Fig. 12a) is correct. If the singular points of the Borel transform move to the positive half-axis as the parameters are varied, then the problem of "non-Borel-summability" should be dealt with. If the analyticity with respect to model parameters is to be preserved, then the moving singular points must not cross the contour C. Therefore, the contour should be deformed as shown in Fig. 12b. Accordingly, correct interpretation of the Borel integral requires that only one parameter γ_i in (7.4) is not zero.

Interpretation in the sense of the Cauchy principal value corresponds to the half-sum of integrals over contours C' and C''. Its difference from the correct interpretation is determined by the half-sum of the integrals over contours C_i going around the singular points z_i (see Fig. 12c). The integral over a contour C_i behaves as $\exp(-z_i/g)$. When interpretation in the sense of the Cauchy principal value is used, nonperturbative contributions of this form must be added to the Borel integral. An expression of this kind was discussed in [118] with regard to the quantum mechanical double-well potential problem.

Can one be sure that the same interpretation will be obtained by analytic continuation of the Borel integral with respect to different model parameters? This question is nontrivial in that it cannot be answered positively in the case of analytic continuation with respect to the coupling constant. Indeed, the action $S\{g, \varphi\}$ (in φ^4 and related theories) can be transformed into $S\{\phi\}/g$, and the change $g \longrightarrow ge^{i\psi}$ is equivalent to $S\{\phi\} \longrightarrow S\{\phi\}e^{-i\psi}$. Thus, the pattern of singularities of the Borel transform in the complex plane is rotated by an angle ψ. The singularities move from the positive half-axis into the left half-plane if $\psi > \pi/2$, which is impossible if the convergence of the functional integral is preserved.

In realistic field theories, the set of parameters is strongly restricted. Due to translational invariance and various symmetries the action can contain only the corresponding invariants, while the renormalizability condition requires that analysis be restricted to small powers of fields and their gradients. Since the coefficients of the highest powers of fields are generally associated with the coupling constant, they are not amenable to analytic continuation. The coefficients of the terms that are quadratic in fields cannot be used either, because their variation may be associated with vacuum instability and phase transitions. The remaining possibilities include the coefficients of the intermediate powers of fields (such as φ^3 in φ^4 theory) and the cross terms representing interactions between different fields. Analyticity with respect to these coefficients is preserved in any part of the complex plane by virtue of (a) convergence of the functional integrals defined on a finite-size lattice; (b) possibility of taking infinite-volume limits everywhere except for phase-transition points, where the system can be partitioned into quasi-independent subsystems owing to the finite correlation length; and (c) possibility of elimination of ultraviolet cutoffs in renormalizable models. These considerations are illustrated here by several examples.

7.1. Zero-Dimensional Model

Consider the integral

$$W(g) = \int_{-\infty}^{\infty} d\varphi \exp\{-\varphi^2 - 2\gamma\sqrt{g}\varphi^3 - g\varphi^4\} = g^{-1/2}\int_{-\infty}^{\infty} d\phi \exp\{-S\{\varphi\}/g\}, \qquad (7.5)$$

as a zero-dimensional analog of the functional integral in φ^4 theories containing cubic terms. According to [50], the singular points in the Borel plane correspond to the extrema of the dimensionless action $S\{\phi\}$:

$$S\{\phi_c\} = \frac{1}{32}[-27\gamma^4 + 36\gamma^2 - 8 \pm \gamma(9\gamma^2 - 8)^{3/2}]. \quad (7.6)$$

If $\gamma = 0$, then there exist two saddle points where the values of action are equal, $\phi_c = \pm i/\sqrt{2}$, and a singularity at $z = S\{\phi_c\} = -1/4$ (on the negative half-axis). If $0 < \gamma < \gamma_c$, where $\gamma_c = (8/9)^{1/2} \approx 0.942$, there exists a pair of complex conjugate points. If $\gamma > \gamma_c$, then this pair lies on the positive half-axis (Fig. 13a), which corresponds to "non-Borel-summability." When γ equals γ_c, another minimum of $S\{\phi\}$ appears on the real axis.

When $\gamma = 0$, the contour C in (5.9) is aligned with the positive half-axis. This choice of C can obviously be retained for $\gamma < \gamma_c$ (Fig. 13b). The configuration obtained by making cuts from the singular points in Fig. 13b to infinity along the rays emanating from the origin can be considered as the quadrilateral $A_1A_2A_3A_4$ with vertices A_2 and A_4 at infinity that can be mapped to the unit circle by the Christoffel–Schwarz integral [51, 119] and then to the plane with a cut (Fig. 8d). The latter mapping is defined as

$$z = p\frac{u}{(1-u)^{\beta}}, \quad p = \frac{\beta^{\beta}(1-\beta)^{1-\beta}}{a}, \quad (7.7)$$

where $2\pi\beta$ is the angle between the cuts made along rays and $1/a$ is the distance from the singular points to the origin. Note that (6.4) and (6.9) are special cases of (7.7) corresponding to $\beta = 2$ and $\beta = 1$. The coefficients U_N of the resummed series in (6.3) are expressed as

$$U_0 = B_0,$$
$$U_N = \sum_{K=1}^{N} B_K p^K \frac{\Gamma(N-K+\beta K)}{\Gamma(N-K+1)\Gamma(\beta K)}, \quad N \geq 0. \quad (7.8)$$

Their asymptotic form in the limit of $N \longrightarrow \infty$,

$$U_N = U_\infty N^{-1+\alpha\beta}, \quad U_\infty = \frac{W_\infty}{\Gamma(\alpha\beta)\Gamma(\alpha+b_0)}, \quad (7.9)$$

determines the asymptotic values of the parameters in $W(g)$ in the strong-coupling limit. When $0 < \gamma < \gamma_c$, the pattern of minima of χ^2 is analogous to that corresponding to $\gamma = 0$ (see Fig. 10). The summation results obtained for $\gamma = 0.25$ and 0.75 are presented in Tables 1 and 2, respectively. As in the case of $\gamma = 0$, the accuracy of summation depends on the error of reconstructing the asymptotics, which increases as γ approaches γ_c (cf. Tables 1 and 2). Indeed, the leading correction to (7.9) scales with $N^{-1+\alpha'\beta}$, and the actual large parameter in the expansion is N^{β} (since $\alpha - \alpha' \sim 1$), which is confirmed by an estimate for summation error analogous to (6.27):

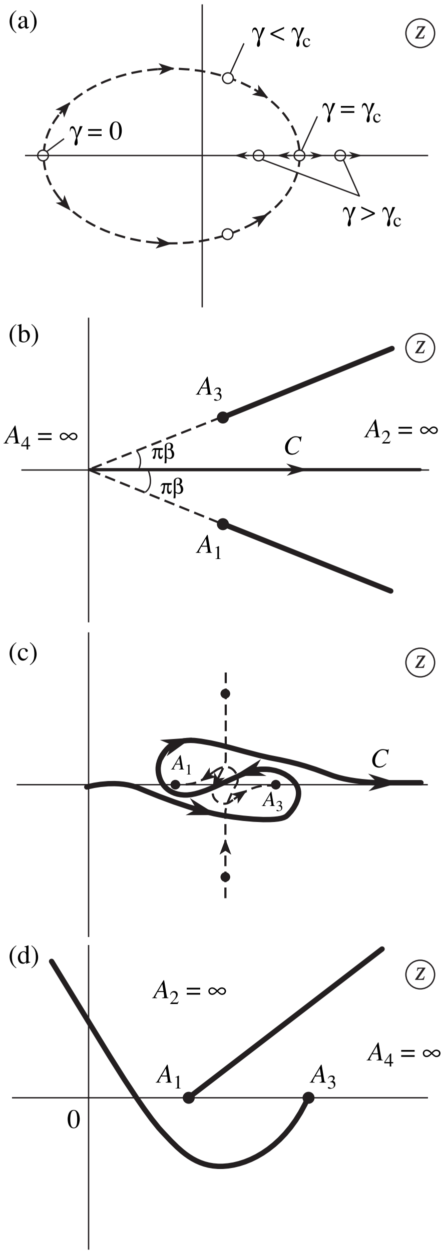


Fig. 13. (a) Singularities in Borel plane for different γ in integral (7.5). Contour of integration: (b) $\gamma < \gamma_c$; (c) $\gamma > \gamma_c$. (d) Point A_1 remains on the physical sheet of the Riemann surface if the cut emanating from A_3 is parabolic.

$$\frac{\delta W(g)}{W(g)} \sim \begin{cases} \Delta, & ag \gtrsim N_c^{\beta}, \\ \Delta\exp\{-(1+\beta)[(1-\beta)^{1-\beta}N_c^{\beta}/ag]^{1/(1+\beta)}\}, & ag \lesssim N_c^{\beta}. \end{cases} \quad (7.10)$$

Table 1. Sum of the series for integral (7.5) with $\gamma = 0.25$

g	$W(g)$	
	exact value	resummed value
0.0625	1.718915	1.718915
0.125	1.674422	1.674422
0.25	1.604821	1.604821
0.50	1.508008	1.508008
1	1.387746	1.387745
2	1.252226	1.252220
4	1.110955	1.11093
8	0.972181	0.97212
32	0.722937	0.72272
$g \longrightarrow \infty$	$1.812g^{-0.25}$	$1.835g^{-0.252}$

Table 2. Sum of the series for integral (7.5) with $\gamma = 0.75$

g	$W(g)$	
	exact value	resummed value
0.0625	1.902930	1.902928
0.125	1.937627	1.93755
0.25	1.903621	1.90300
0.50	1.787743	1.7851
1	1.615170	1.608
2	1.419861	1.406
4	1.226524	1.205
8	1.048303	1.020
32	0.753306	0.714
$g \longrightarrow \infty$	$1.812g^{-0.25}$	$1.885g^{-0.275}$

The parameter N^{β} decreases as $\gamma \longrightarrow \gamma_c$, since $\beta \longrightarrow 0$ and the value of N cannot be increased indefinitely for technical reasons. Thus, the algorithm formulated above cannot be used to deal with the case of $\gamma = \gamma_c$, even though approaching this state to an arbitrarily small distance does not contradict any fundamental principle.

To describe the case of $\gamma > \gamma_c$, the degeneracy of the singular points at $\gamma = \gamma_c$ is eliminated by adding a constant imaginary quantity $i\delta$ to γ. When $\gamma \approx \gamma_c$, saddle-point action (7.6) contains the singular contribution $(\gamma - \gamma_c)^{3/2}$. As γ increases, vertical displacement of the singular points is followed by horizontal after the rotation by an angle of $3\pi/2$ is performed, and the contour C folds (Fig. 13c). After the cut emanating from A_3 is rotated so as to coincide with the positive half-axis, point A_1 lies on a different sheet of the Riemann surface and does not contribute to the divergence of the perturbation series, which can be verified directly by calculating the expansion coefficients. This agrees with the fact that the formally calculated contribution of A_1 to the coefficients is purely imaginary. To hold the singular point A_1 on the physical sheet of the Riemann surface, the cut emanating from A_3 must be curved. If the cut is parabolic (see Fig. 13d), the following constructive algorithm can be used: the mapping $w = \sqrt{z - A_1}$ transforms the plane in Fig. 13d into a plane with a straight cut, which is mapped to a unit circle by the Christoffel–Schwarz integral, and then any desired domain can be obtained (see Fig. 8c).

7.2. Double-Well Potential

Consider the ground state of a quantum particle in the potential

$$U(x) = \frac{1}{2}x^2 - \gamma\sqrt{g}x^3 + \frac{1}{2}gx^4, \qquad (7.11)$$

which reduces to anharmonic oscillator (1.1) when $\gamma = 0$ and becomes a double-well potential with symmetric minima when $\gamma = 1$. The latter model is of interest as a typical case of two degenerate vacuums. It was asserted in [116] that problems of this type cannot be solved by summation of perturbation series in principle. Model (7.11) reduces to a one-dimensional field theory with two instantons when $0 < \gamma < 1$, and the corresponding dimensionless action is

$$S\{\phi_c\} = -\frac{2}{3} + \gamma^2 - \frac{1}{2}\gamma(\gamma^2 - 1)\left[\ln\frac{1+\gamma}{1-\gamma} \pm \pi i\right]. \qquad (7.12)$$

If m-instanton configurations are taken into account, then the singularities of the Borel transform lie at the points $z_m = -(2/3)m$ on the negative half-axis when $\gamma = 0$ and on two rays emanating from the origin when $0 < \gamma < 1$ (see Fig. 14a). In the latter case, the perturbation series can be resummed by using the conformal mapping defined by (7.7). The case of $\gamma = 1$ corresponds to the critical state approached as $\beta \longrightarrow 0$. This state is unreachable in a rigorous sense, but there is no fundamental principle that forbids approaching it to an arbitrarily small distance. This can be done without using perturbation series in terms of an arbitrary γ. It will suffice to analyze the change in the Lipatov asymptotic form caused by a small deviation of γ from unity [43].

7.3. Yang–Mills Theory

In the Yang–Mills theories, the Borel integral can be interpreted by using a procedure that resembles analytical continuation with respect to the coupling between fields, but preserves gauge invariance. This is facilitated by invoking the results obtained in [18], where an SU(2) Yang–Mills field A^a_ν coupled to a complex scalar field φ was considered,

$$S\{A,\varphi\} = \int d^4x \left\{ \frac{1}{4}(F^a_{\mu\nu})^2 + \left|(\partial_\mu - ig\tau^a A^a_\mu)\varphi\right|^2 + \frac{1}{2}\lambda^2|\varphi|^4 \right\}, \tag{7.13}$$

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g\epsilon_{abc}A^b_\mu A^c_\nu,$$

with $\tau^a = \sigma^a/2$ and σ^a denoting Pauli matrices. After changing to new field variables, $A \longrightarrow B/g$, $\varphi \longrightarrow \phi/\lambda$, action (7.13) is represented as

$$S\{A,\varphi\} = \frac{S_0\{B\}}{g^2} + \frac{S_1\{B,\phi\}}{\lambda^2} \equiv \frac{S\{B,\phi\}}{g^2}, \tag{7.14}$$

where the last equality is written by introducing $\chi = \lambda^2/g^2$. In this theory, an arbitrary functional $Z(g^2, \lambda^2)$ can be represented as a double series in powers of g^2 and λ^2,

$$Z(g^2,\lambda^2) = \sum_{K,M} Z_{K,M}g^{2M}\lambda^{2K}, \tag{7.15}$$

with coefficients $Z_{K,M}$ determined by the saddle-point configurations of functional (7.14) modified by adding $-M\ln g^2$ and $-K\ln\lambda^2$. The saddle-point values of g^2 and λ^2 are

$$g^2_c = \frac{S_0\{B_c\}}{M}, \quad \lambda^2_c = \frac{S_1\{B_c,\phi_c\}}{K}, \tag{7.16}$$

while the saddle-point field configuration is given by

$$\begin{aligned} B^a_\mu(x) &= 4\eta^a_{\mu\nu}x_\nu\frac{\rho^4-1}{(x^2+\rho^2)(\rho^2x^2+1)}, \\ \phi(x) &= \pm i\sqrt{\chi}U\frac{4\sqrt{3}}{[(x^2+\rho^2)(\rho^2x^2+1)]^{1/2}}, \\ \rho^4 &= 12\chi - 1, \end{aligned} \tag{7.17}$$

where $\eta^a_{\mu\nu}$ denotes 't Hooft matrices, U is a constant spinor ($UU^* = 1$). The saddle-point action is expressed as

$$\begin{aligned} S\{B_c,\phi_c\} &= 16\pi^2\left[-2 + \frac{3(\sinh 4\xi_0 - 4\xi_0)}{2\sinh^2 2\xi_0}\right], \\ S_1\{B_c,\phi_c\} & \\ &= 16\pi^2\chi(-6\cosh 2\xi_0)e^{-2\xi_0}\frac{2\xi_0\cosh 2\xi_0 - \sinh 2\xi_0}{\sinh^3 2\xi_0}, \\ e^{2\xi_0} &= \rho^2, \end{aligned} \tag{7.18}$$

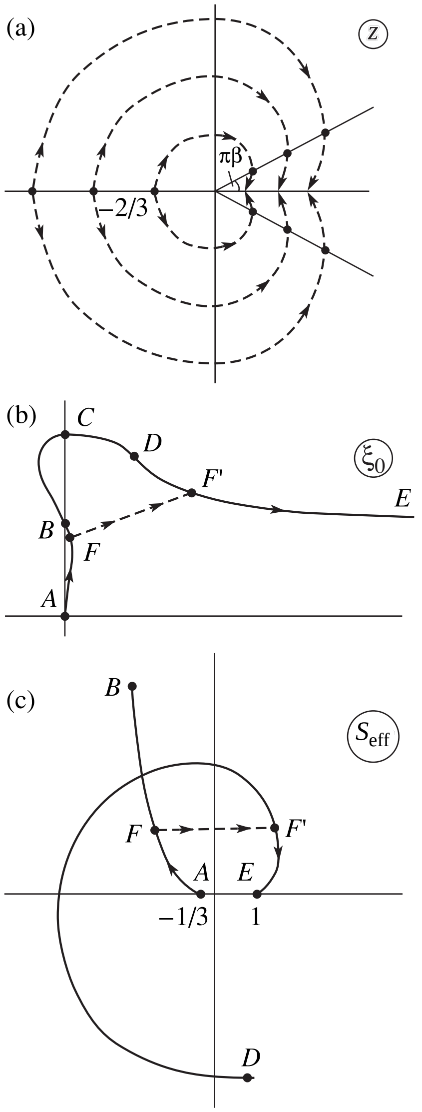


Fig. 14. (a) Singularities in the Borel plane corresponding to different γ in potential (7.11). Interpretation of the Borel integral in the Yang–Mills theory: (b) curve *ABCDE* in the ξ_0 plane defined by the condition $\mathrm{Im}f(\xi_0) = 0$ and (c) the corresponding action S_{eff} (in units of $16\pi^2$).

while the corresponding value is

$$\begin{aligned} \chi &= \frac{\lambda^2_c}{g^2_c} = \frac{M}{K}\frac{S_1\{B_c,\phi_c\}}{S_0\{B_c\}} \\ \text{or}\quad \frac{\chi S\{B_c,\phi_c\}}{S_1\{B_c,\phi_c\}} &= \frac{M+K}{K}. \end{aligned} \tag{7.19}$$

If K and M in (7.15) satisfy some constant relations,

$$K = \gamma N, \quad M = \delta N, \quad \gamma + \delta = 1, \tag{7.20}$$

then the asymptotic behavior of the expansion coefficients is described by the formula

$$Z_{K,M} = Z_{\gamma N,\,\delta N} \propto N!\,\mathrm{Re}[S_{\mathrm{eff}}(\xi_0)]^{-N}, \tag{7.21}$$

where

$$S_{\mathrm{eff}}(\xi_0) = 16\pi^2\left[-2 + \frac{3(\sinh 4\xi_0 - 4\xi_0)}{2\sinh^2 2\xi_0}\right] \times \left(\frac{e^{4\xi_0}+1}{12}\right)^{\gamma}, \tag{7.22}$$

and ξ_0 is found by solving the second equation in (7.19) rewritten by using (7.18) as

$$f(\xi_0) = e^{2\xi_0}\tanh 2\xi_0 \times \frac{2\xi_0 + \frac{1}{3}\sinh^2 2\xi_0 - \frac{1}{2}\sinh 2\xi_0 \cosh 2\xi_0}{2\xi_0\cosh 2\xi_0 - \sinh 2\xi_0} = \frac{1}{\gamma}. \tag{7.23}$$

Equation (7.23) has a pair of complex conjugate solutions, and the symbol Re in (7.21) is results from summation over these solutions. Figure 14b shows the curve *ABCDE* defined by the condition $\mathrm{Im} f(\xi_0) = 0$. On this curve, $\mathrm{Re} f(\xi_0)$ varies from 1 to ∞ along segment *AB*, from –∞ to 0 along *BC*, and from 0 to ∞ along *CE*. Physical values of γ ($0 \le \gamma \le 1$) correspond to segments *AB* and *DE*, where $|S_{\mathrm{eff}}(\xi_0)|/16\pi^2$ varies from 1/3 to 5.4 and from 4.2 to 1, respectively (see Fig. 14c). Asymptotic form (7.21) is determined by the saddle point with the minimum value $|S_{\mathrm{eff}}(\xi_0)|$, and the variation of γ from 1 to 0 corresponds to the movement of a point along the trajectory *AFF'E* with a jump between points *F* and *F'*, which are associated with equal values of γ and $|S_{\mathrm{eff}}(\xi_0)|$. The jump in action (see Fig. 14c) can be eliminated by moving via complex values of γ defined by relation (7.23) on the segment *FF'* in the complex ξ_0 plane (see Fig. 14b).

Point *A* corresponds to the value $\rho^2 = 1$ for which the Yang–Mills field vanishes (see (7.17)), and $S_{\mathrm{eff}}(\xi_0)$ corresponds to the saddle-point action in φ^4 theory. Conversely, the parameter χ increases indefinitely at the right endpoint of the curve ($\xi_0 \longrightarrow \infty + i\pi/4$), the field φ vanishes accordingly, and $S_{\mathrm{eff}}(\xi_0)$ corresponds to the value of action for an instanton–anti-instanton pair for the pure Yang–Mills theory. If coefficients (7.21) formally define the series

$$Z(\tilde{g}^2) = \sum_N Z_{\gamma N,\,\delta N}\tilde{g}^N, \quad \tilde{g} = g^{2\gamma}\lambda^{2\delta}, \tag{7.24}$$

then γ can be varied to perform a continuous change from the series for φ^4 theory to the series for the Yang–Mills theory and to monitor the evolution of the singularities of the Borel transform at $z = S_{\mathrm{eff}}(\xi_0)$ and $z = S^*_{\mathrm{eff}}(\xi_0)$. If the *m*-instanton configurations are taken into account, then the result is similar to that obtained for the double-well potential (Fig. 14a). Therefore, the summation of the series for the Yang–Mills theory should make use of conformal mapping (7.7) with an extremely small parameter β.

8. GELL-MANN–LOW FUNCTIONS IN BASIC FIELD THEORIES

This section presents a scheme for finding the Gell-Mann–Low functions in basic field theories with arbitrary coupling constants (see Fig. 15).

8.1. φ^4 Theory

The first attempt to recover the Gell-Mann–Low function in the four-dimensional scalar φ^4 theory was made in [11]. The analysis of the strong-coupling limit presented in [12] predicted the asymptotic behavior $0.9g^2$, which differs from the one-loop result $1.5g^2$ valid for $g \longrightarrow 0$ only by a numerical factor. Similar asymptotic behavior, $1.06g^{1.9}$, was obtained in [13]. The variational perturbation theory developed in [120] predicts $2.99g^{1.5}$. All of these results indicate that φ^4 theory is non-self-consistent (or "trivial"), which contradicts the absence of renormalon singularities established in Section 5. An additional argument follows from the fact that φ^4 theory can be rigorously derived from a reasonable model of a disordered system [103, 121–123], which is well defined in the continuum limit.

The Gell-Mann–Low function can be found by means of the algorithm described in Section 6.2 with $\beta(g)$ playing the role of $W(g)$ [65]. The input data used here are the same as in [12]: the values of the first four coefficients of the β function expansion in the subtraction scheme developed in [124, 125],

$$\beta(g) = \frac{3}{2}g^2 - \frac{17}{6}g^3 + \frac{154.14}{8}g^4 - \frac{2338}{16}g^5 + \dots, \tag{8.1}$$

and their high-order asymptotic form [7] with a correction calculated in [126],

$$\beta_N = \frac{1.096}{16\pi^2}N^{7/2}N!\left\{1 - \frac{4.7}{N} + \dots\right\}. \tag{8.2}$$

This asymptotic expression is determined by the expansion coefficients for the invariant charge, which corresponds here to the vertex with $M = 4$ (cf. (4.1.18)). The "natural" normalization is used for charge g, with the parameter a in (2.5) set equal to unity. In this case, the nearest singular point of the Borel transform is sepa-

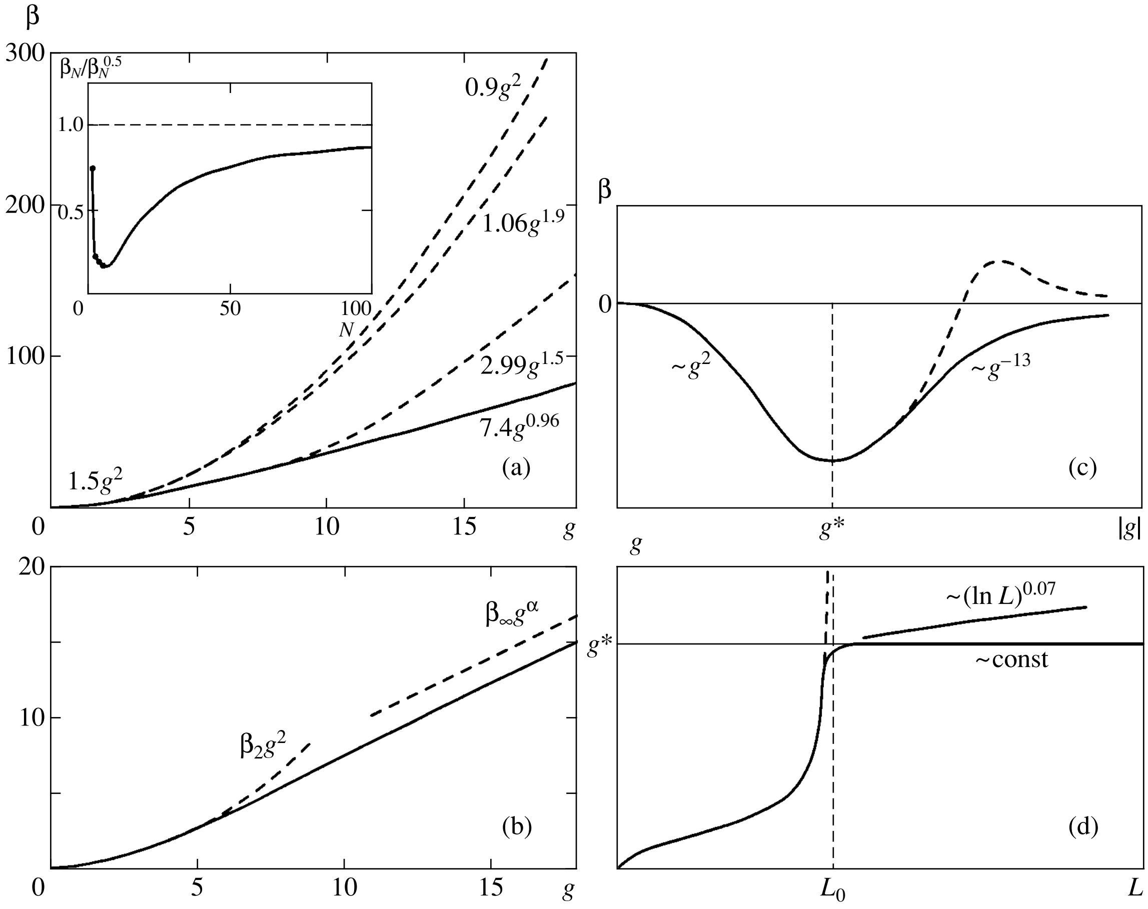


Fig. 15. Gell-Mann–Low functions for (a) φ^4 theory, (b) QED, and (c) QCD. (d) Effective coupling in QCD.

rated by unit distance from the origin, and the characteristic scale of variation of β function is $g \sim 1$.

Interpolation was performed using (6.24) with $\tilde{b} = 4$ as an optimal value. Coarse optimization of χ^2 as a function of $\tilde{N}$ was performed for several constant values of b_0 [65] to determine the range of interpolations ($\tilde{N}$ between –0.5 and 0.5) consistent with the power-law asymptotic behavior of $W(g)$. Figure 16 shows the behavior of U_N for a nearly optimal interpolation with $\tilde{N} = 0$. Since all curves except for those corresponding to $b_0 \gg 1$ and $b_0 \approx -2$ (which approach their respective asymptotes at slower rates) tend to constants at large N, the value of α is close to unity. This result agrees with the value of α_{eff} at the right-hand minimum of χ^2, the location of the left-hand minimum of χ^2, and the behavior of U_∞ near its zero (see Fig. 17). Figure 18 illustrates the dependence of the results on $\tilde{N}$. The behavior of α corresponding to the "ideal" situation shown in Fig. 11b is obtained by widening the error corridor by a factor of 2 (short-dashed curves in Fig. 18a). The resulting value $\alpha = 0.96$ is consistent with all results obtained for various $\tilde{N}$. The "ideal" situation for W_∞ is obtained immediately (Fig. 18b), and the corresponding value $W_\infty = 7.4$ is consistent with all results. Thus,

$$\alpha = 0.96 \pm 0.01, \quad W_\infty = 7.4 \pm 0.4. \tag{8.3}$$

A similar pattern is observed when $\tilde{b}$ is varied in (6.24) [65].

Figure 15a compares the β function obtained for $g \le 20$ by series summation (solid curve) with results obtained in [12, 13, 120] (upper, middle, and lower dashed curves, respectively). The asymptotic form of $\beta(g)$ found in [12, 13] corresponds to the stable line segment $\tilde{U}_N \approx 1.1N$ at $N \le 10$ in Fig. 16, which is inevitably interpreted as the true asymptote if it is calculated by using only the known expansion coefficients. Actually, this segment is associated with a dip in the coefficient function at $N \lesssim 10$ (see insert to Fig. 15a), which manifests itself in the behavior of the β function by a one-loop law[17] with domain extending to $g \sim 10$ [65]. Thus, the results obtained in [12, 13] reflect certain

---

[17] Being more pronounced for the Borel transform, this behavior is less obvious for the β function because of the integration in (6.2).

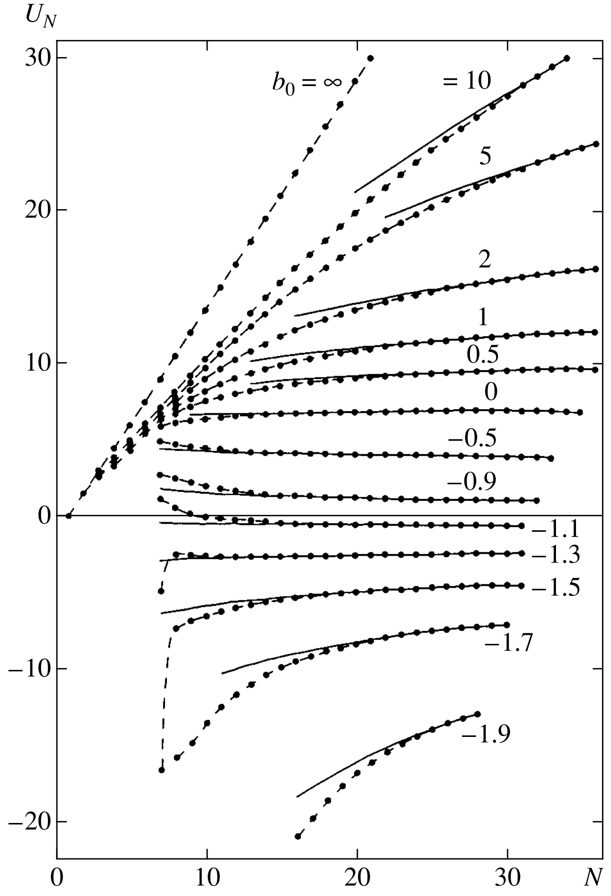


Fig. 16. Coefficients $\tilde{U}_N$ versus N for different b_0 (symbols) and corresponding approximate power laws (solid curves) obtained for φ^4 theory by nearly optimal interpolation with $\tilde{b} = 4$ and $\tilde{N} = 0$.

properties of the β function and should not be interpreted as essentially incorrect (see detailed discussion in [65, Section 8.3]). The variational perturbation theory developed in [120] provides a somewhat better description of the region of $g \lesssim 10$ in Fig. 15a, but does not guarantee correct results in the strong-coupling limit even theoretically.

The value of α obtained is close to unity. Even though the deviation from unity exceeds the error, the exact equality $\alpha = 1$ cannot be ruled out, because asymptotic expansion (6.7) may contain logarithmic corrections,

$$W(g) = W_\infty g^\alpha (\ln g)^{-\gamma}, \quad g \longrightarrow \infty, \qquad (8.4)$$

which may be interpreted as a slight decrease in α if $\gamma < 0$. In this case, expansion (6.11) contains the factor $(\ln N)^{-\gamma}$, while U_∞ does not change, and the resulting U_N can be fitted by using (8.4) with

$$\alpha = 1, \quad \gamma \approx 0.14, \quad W_\infty \approx 7.7 \qquad (8.5)$$

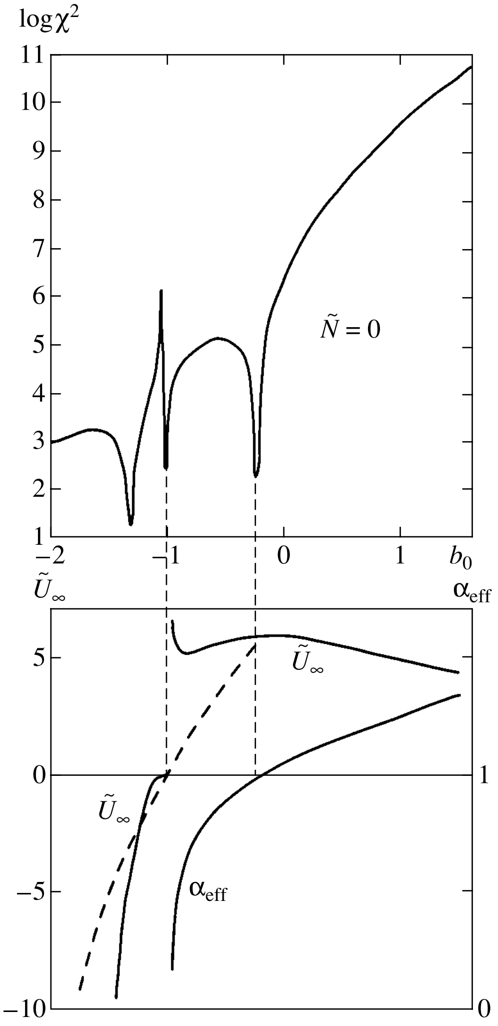


Fig. 17. Pattern of χ^2 minima obtained for φ^4 theory by using the interval $20 \le N \le 40$ and curves of $\alpha_{\rm eff}$ and $\tilde{U}_\infty$ versus b_0 calculated by interpolation with $\tilde{N} = 0$. Dashed curve is $U_\infty(b_0)$ corresponding to constant $\alpha = 1$.

without any increase in χ^2. Logarithmic branching appears to be quite plausible for the following reasons.

1. Logarithmic branching is inevitable when the exact equality $\alpha = 1$ holds. Indeed, series (6.1) can be represented as the Sommerfeld–Watson integral [7, 13]

$$W(g) = \sum_{N = N_0}^{\infty} W_N(-g)^N = -\frac{1}{2i}\oint_C dz \frac{\mathcal{W}(z)}{\sin \pi z} g^z, \qquad (8.6)$$

where $\mathcal{W}(z)$ is the analytic continuation of W_N to the complex plane ($\mathcal{W}(N) = W_N$), C is a contour encompassing the points N_0, $N_0 + 1$, $N_0 + 2$, …. When g is large, the contour C can be extended and shifted leftwards until the rightmost singular point of $\mathcal{W}(z)/\sin\pi z$ at $z = \alpha$ is reached. This singularity determines the behavior of $W(g)$ as $g \longrightarrow \infty$. Power law (6.7) and asymptotic formula (8.4) correspond to the existence of a simple pole at $z = \alpha$ and a singularity of the form $(z-\alpha)^{\gamma-1}$, respectively.

The term β_0 in expansion (8.1) vanishes by definition. However, the zero value of the coefficient β_1 fortuitous: in the $(4-\epsilon)$-dimensional φ^4 theory, it has a finite value on the order of ϵ, and $\mathcal{W}(1) \sim \epsilon$ accordingly. As $\epsilon \longrightarrow 0$, the four-dimensional limit value $\mathcal{W}(1) = 0$ is obtained, and there is no simple pole if $\alpha = 1$. If zero is approached according to the law $\mathcal{W}(z) = \omega_0(z-1)^\gamma$, then

$$\beta(g) = \frac{\omega_0}{\Gamma(1-\gamma)} g(\ln g)^{-\gamma}, \quad g \longrightarrow \infty, \tag{8.7}$$

and the positive value of γ has a natural explanation.

2. The class of field theories with nonlinear terms φ^n (generalizations of φ^4 theory) analyzed in [127] for spaces of dimension $d = 2n/(n-2)$ is characterized by logarithmic behavior. In theories of this type, the coefficient β_1 vanishes, but becomes finite as d decreases. Therefore, $\mathcal{W}(1) = 0$ as shown above. The Gell-Mann–Low function can be calculated exactly as $n \longrightarrow \infty$ [127], and the rightmost singularity of $\mathcal{W}(z)$ has the form $(z-1)^{3/2}$, which corresponds to the asymptotic behavior $\beta(g) \propto g(\ln g)^{-3/2}$. By continuity, nonanalyticity of the form $(z-1)^\gamma$ should hold for finite n, and the singularity at $z = 1$ should remain rightmost. Therefore, asymptotic behavior (8.7) is natural for field theories of this kind, and it is no surprise that it holds even for $n = 4$. Note that W_∞ is negative as $n \longrightarrow \infty$, and the Gell-Mann–Low function has a zero. A similar conclusion can be drawn for φ^4 theory by straightforward extrapolation to $n = 4$ [127]. Actually, this extrapolation should use the fact that the exponent γ varies from 3/2 to small values (see (8.5)). Accordingly, the asymptotic expression obviously changes sign when $\gamma = 1$ according to (8.7) (ω_0 is positive since $\mathcal{W}(2) \sim \omega_0$ and β_2 is positive [127]).

Thus, one has to choose between two possibilities: power law (6.7) with α slightly below unity and asymptotic expression (8.7) with $\gamma > 0$. In either case, φ^4 theory is self-consistent, which contradicts the widespread view that φ^4 theory is trivial. Let us discuss the origin of this belief (for a more detailed discussion, see [65, Section 8.4]).

It has been rigorously proved that φ^4 theory is trivial for $d > 4$ and nontrivial for $d < 4$ [128, 129]. The inequalities obtained for $d = 4$ are "just a bit" insufficient for proving triviality [130, Section 14]. For mathematicians, this is an annoying minor problem, and the triviality of φ^4 theory is commonly regarded as "almost proved." For physicists, there is no reason to be so optimistic about it: from the modern perspective, the afore-mentioned results obtained for $d \neq 4$ are nothing more than elementary corollaries of renormalizability theory and one-loop renormalization group analysis. However, the physics of the theory with $d = 4$ are extremely complicated, and no analytical approach to the problem has been found to this day.

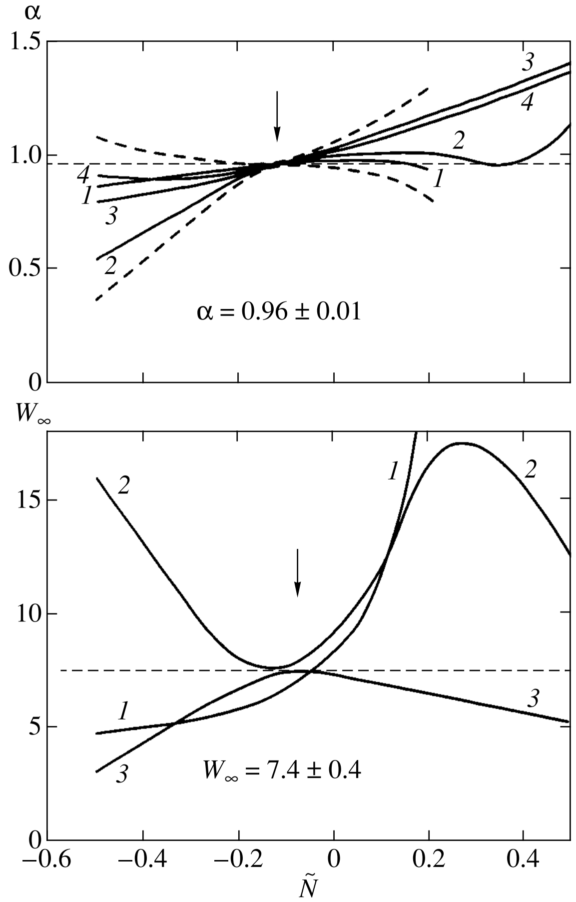


Fig. 18. Curves of α and W_∞ estimated for φ^4 theory: numbers at curves correspond to estimates in Section 6.2; short-dashed curves illustrate the widening of the error corridor for α by a factor of 2.

It is generally believed that the triviality of φ^4 theory can be demonstrated by numerical experiments on lattices. However, most of them reveal only a decrease in the effective charge $g(L)$ with increasing L, which is quite natural because the β function has no zero, whereas convincing evidence of "zero charge" can hardly be obtained on any finite-size lattice. There is considerable misunderstanding with regard to charge normalization. Even under the "natural normalization" used here, a quadratic law holds for $g \sim 10$. Under conventional normalization conditions, it holds on wider intervals, for example, even for $g \sim 2000$ when the coupling term is written as $g\varphi^4/4!$. Accordingly, behavior of any variable is impossible to distinguish from "trivial" over a wide range of parameter values. The very concept of triviality is frequently misunderstood. Many authors relate it to the mean-field values of critical exponents in the four-dimensional theory of phase transitions, but this indisputable fact is due to the mere absence of nontrivial zero of the β function.

Issues related to triviality were analyzed by Agodi, Consoli, and others in a recent series of publications (e.g., see [131, 132]). An unconventional scenario for continuum limit in φ^4 theory was proposed and claimed to be logically consistent. The validity of the conventional perturbation theory was basically denied, which seems to be a premature conclusion. Since the numerical results used as supportive evidence were obtained in computations on lattices in the weak-coupling limit, they cannot provide any information about triviality. The analyses presented in [131, 132] were performed to resolve the difficulties arising in the Higgs sector of the Standard Model in view of the triviality of φ^4 theory. No issues of this kind arise when the theory is self-consistent.

8.2. Quantum Electrodynamics

In QED, four terms of the expansion of the β function are known in the MOM scheme [133]:

$$\begin{aligned}\beta(g) &= \frac{4}{3}g^2 + 4g^3 + \left[\frac{64}{3}\zeta(3) - \frac{202}{9}\right]g^4 \\ &+ \left[186 + \frac{256}{3}\zeta(3) - \frac{1280}{3}\zeta(5)\right]g^5 + \ldots,\end{aligned} \tag{8.8}$$

and the corresponding asymptotic expression is

$$\beta_N^{\text{as}} = \text{const} \times 4.886^{-N}\Gamma\left(\frac{N+12}{2}\right), \quad N \longrightarrow \infty. \tag{8.9}$$

It is identical, up to a constant factor, to the asymptotic behavior of coefficients for the invariant charge [7], which depends on gD in QED, where D is the photon propagator (see (4.2.11) for $M = 2$ and $L = 0$).

The summation procedure for this series must be modified, as compared to that developed in Section 6, because the Lipatov asymptotic form is $ca^N\Gamma(N/2 + b)$ rather than $ca^N\Gamma(N + b)$. The Borel transform is

$$\begin{aligned}\beta(g) &= \int_0^\infty dx e^{-x} x^{b_0 - 1} B(ag\sqrt{x}), \\ B(z) &= \sum_{N=0}^\infty B_N(-z)^N, \\ B_N &= \frac{\beta_N}{a^N\Gamma(N/2 + b_0)},\end{aligned} \tag{8.10}$$

where b_0 is an arbitrary parameter. The conformal mapping $z = u/(1 - u)$ is applied to obtain a convergent series in u for the Borel transform, with coefficients

$$\begin{aligned}U_N &= \sum_{K=1}^N B_K(-1)^K C_{N-1}^{K-1} \quad (N \geq 1), \\ U_0 &= B_0,\end{aligned} \tag{8.11}$$

whose behavior at large N,

$$U_N = U_\infty N^{\alpha - 1}, \quad U_\infty = \frac{\beta_\infty}{a^\alpha\Gamma(\alpha)\Gamma(b_0 + \alpha/2)}, \tag{8.12}$$

determines the parameters of the asymptotic expression $\beta(g) = \beta_\infty g^\alpha$ as $g \longrightarrow \infty$.

Interpolation was performed by using (6.24) with $\tilde{b} = b - 1/2 = 5.5$ [65]. In contrast to φ^4 theory, the constant factor in (8.9) is not known. Technically, this is not a problem, because the constant c can be factored into the curly brackets in (6.24) to replace 1 with a parameter $\tilde{A}_0$ treated as unknown and calculated by interpolation. However, this leads to a much higher uncertainty in the normalized coefficient function $F_N = \beta_N/\beta_N^{\text{as}}$: its values $F_2 = 63.1$, $F_3 = -7.02$, $F_4 = 0.34$, and $F_5 = 1.23$ (measured in units of 10^{-3}) exhibit only weak convergence to a constant, and the predicted $\tilde{A}_0 = \lim_{N \to \infty} F_N$ varies by orders of magnitude as a function of $\tilde{N}$. Nevertheless, the "superstability" of the algorithm mentioned above (see Section 6.2) suggests that reasonable results can be obtained even in this situation. To verify this possibility, a test experiment was performed for φ^4 theory. The complete input data (including both coefficients β_2, β_3, β_4, and β_5 and parameters $\tilde{A}_0$ and $\tilde{A}_1$) were used to obtain $\alpha = 0.96 \pm 0.01$ and $\beta_\infty = 7.4 \pm 0.4$ (recall Section 8.1). Similar computations performed without using $\tilde{A}_0$ and $\tilde{A}_1$ resulted in $\alpha = 1.02 \pm 0.03$ and $\beta_\infty = 1.7 \pm 0.3$. Since the uncertainty in the coefficient function (estimated by varying $\tilde{N}$ within unity about its optimal value) is a few percent in the former case and more than an order of magnitude in the latter, this robustness of results is quite satisfactory.[18] It is clear that the results presented below should be considered as a zeroth approximation.

Coarse optimization of χ^2 as a function of $\tilde{N}$ was performed to determine the range of interpolations ($-0.5 \lesssim \tilde{N} \lesssim 1.0$) for which U_N may exhibit power-law behavior. The dependence of χ^2 and U_∞ on α and b_0 illustrated by Fig. 19 implies that $\alpha \approx 1$. Indeed, U_∞

---

[18] The shift in β_∞ cannot be controlled by error estimation. This is obviously explained by the fact that the procedure of error estimation validated in [65] is justified only when the discrepancy with the exact result is so small that all deviations can be linearized.

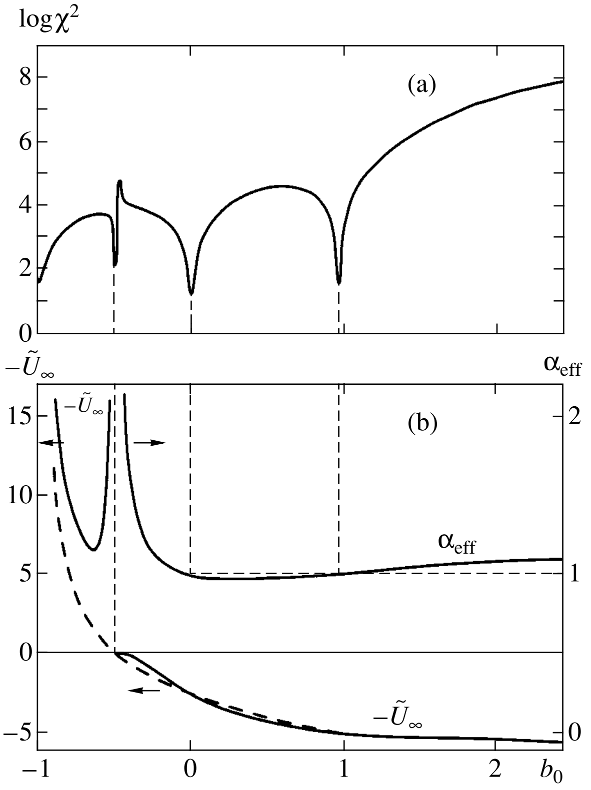


Fig. 19. Pattern of χ^2 minima and curves of $\alpha_{\rm eff}$ and $\tilde{U}_\infty$ versus b_0 obtained for quantum electrodynamics by using the interval $20 \le N \le 40$ (notation as in Fig. 17).

changes sign when $\beta_0 = -\alpha/2 \approx -0.5$ (see (8.12)). The same value of b_0 corresponds to the minimum of χ^2 due to zero value of the leading contribution to the asymptotic $U_\infty N^{\alpha-1}$. The values of $\alpha_{\rm eff}$ corresponding to the minima of χ^2 at $b_0 = -\alpha'/2, -\alpha''/2, \ldots$, where the corresponding corrections to (8.12) vanish, are closest to the exact value $\alpha \approx 1$.[19]

Figures 20a and 20b show several estimates for α and β_∞ as functions of $\tilde{N}$. The values of α obtained for $\tilde{N} \le 0.25$ are consistent with a value slightly below unity. The systematic growth to 1.08 observed at $\tilde{N} > 0.25$ cannot be controlled by error estimation, but the corresponding minima of χ^2 are weak and unstable. Similar behavior is characteristic of β_∞. The results obtained for the central part of the examined interval of $\tilde{N}$ are accepted as more reliable, with a conservative error estimate including systematic variations:

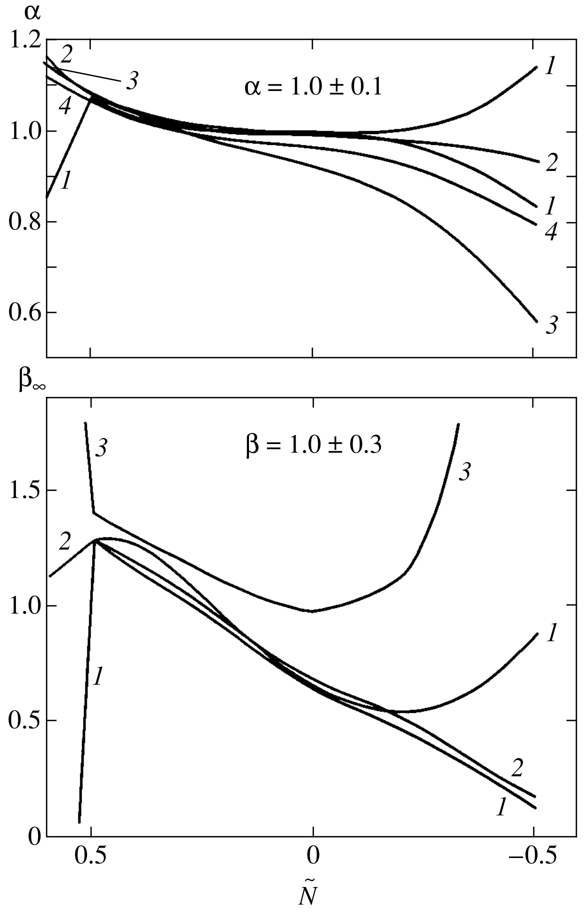


Fig. 20. Curves of α and W_∞ estimated for QED: numbers at curves correspond to estimates in Section 6.2.

$$\alpha = 1.0 \pm 0.1, \quad \beta_\infty = 1.0 \pm 0.3. \qquad (8.13)$$

In view of the above remarks concerning errors, even this estimate is somewhat unreliable.

Figure 15b shows the results obtained by summing the series for $\tilde{N} = 0.2$ and $b_0 = 0$. The one-loop law $\beta_2 g^2$ is matched with the asymptotic $\beta_\infty g^\alpha$ at $g \sim 10$. The difference between $\beta(g)$ and the one-loop result is negligible at $g < 5$. The asymptotic $\beta(g)$ agrees with the upper bound in the inequality $0 \le \beta(g) < g$ derived in [134] from a spectral analysis within the estimated error. If $\alpha = 1$ and $\beta_\infty = 1$, then the fine structure constant in pure electrodynamics increases as L^{-2} in the small-length limit.

Even specialists disagree on the interpretation of the results obtained in lattice QED [135, 136]. Overall, they point to the triviality in Wilson's sense: the β function does not have a nontrivial zero, and phase transitions are characterized by mean-field critical exponents. This conclusion agrees with the results presented above.

---

[19] Usually, only the minima of χ^2 corresponding to α and α' (recall Section 6.2) are observed in test examples. Additional minima may appear when certain relations between the coefficients W'_∞, W''_∞, … are satisfied. This must occur when the amount of available information is small, as in the test computations on φ^4 theory described here.

8.3. QCD

In QCD, the first four terms of the expansion of the Gell-Mann–Low function are known in the MS scheme [137]:

$$\beta(g) = -\sum_{N=0}^{\infty}\beta_N g^N = -\beta_2 g^2 - \beta_3 g^3 - \beta_4 g^4 - \ldots, \tag{8.14}$$

$$g = \frac{\bar{g}^2}{16\pi^2},$$

$$\beta_2 = 11 - \frac{2}{3}N_f, \quad \beta_3 = 102 - \frac{38}{3}N_f,$$

$$\beta_4 = \frac{2857}{2} - \frac{5033}{18}N_f + \frac{325}{54}N_f^2, \tag{8.15}$$

$$\beta_5 = \left[\frac{149753}{6} + 3564\zeta(3)\right] - \left[\frac{1078361}{162} + \frac{6508}{27}\zeta(3)\right]N_f + \left[\frac{50065}{162} + \frac{6472}{81}\zeta(3)\right]N_f^2 + \frac{1093}{729}N_f^3,$$

where $\bar{g}$ is the coupling constant in QCD Lagrangian (4.5.7). The asymptotic form of the coefficients in series (8.14)

$$\beta_N = \mathrm{const}\,\Gamma\left(N + 4N_c + \frac{11(N_c - N_f)}{6}\right) \tag{8.16}$$

is determined by the expansion of the invariant charge, which can be found by using any vertex in view of the generalized Ward identities [47]. Formula (8.16) for $N_c = 2$ and $N_f = 0$ agree with the result obtained in [17].

Since (8.14) is a nonalternating series, its summation should be performed by the method described in Section 7. However, a simpler procedure [47] can be applied by assuming that $B(z) \sim z^{\alpha}$ at infinity. Irrespective of interpretation of Borel integral (7.4), the result is

$$\begin{aligned} \beta(g) &= \beta_\infty g^\alpha, \quad g \longrightarrow \infty, \\ \beta(g) &= \bar{\beta}_\infty |g|^\alpha, \quad g \longrightarrow -\infty, \end{aligned} \tag{8.17}$$

where the exact relation between β_∞ and $\bar{\beta}_\infty$ depends on γ_i and C_i. However, since $\beta_\infty \sim \bar{\beta}_\infty$ in the general case, the summation of series (8.14) for negative g can be used to determine the exponent α and estimate β_∞.

Interpolation of the coefficient function is performed by using (6.24) with $\tilde{b} = b - 1/2$. As in QED, the parameter c in the Lipatov asymptotic form is not known. In Section 8.2, it was calculated in the course of interpolation. In the present case, the results of an analogous calculation are characterized by considerable uncertainties, which cannot be reduced by optimization. For this reason, interpolation was performed by starting from a trial value of c and then varying it between 10^{-5} and 1.[20] The change in the results due to the variation was negligible as compared to other uncertainties. The results presented below were obtained for $N_c = 3$, $N_f = 0$, and $c = 10^{-5}$.

By finding a power-law fit for U_N and analyzing χ^2 as a function of $\tilde{N}$ [47], it was found that the minimal values of χ^2 correspond to $0.5 \lesssim \tilde{N} \lesssim 2.0$. Thus, the range of interpolations consistent with the power-law behavior of U_N was determined. The typical curves of χ^2 and effective U_∞ and α plotted versus b_0 in Fig. 21 demonstrate that $\alpha \approx -15$. Indeed, U_∞ changes sign (see (6.12)) at $b_0 = -\alpha \approx 15.5$, and the left-hand minimum of χ^2 is located at the same point. A similar estimate, $\alpha \approx -15$, is obtained by using the value of α_{eff} at

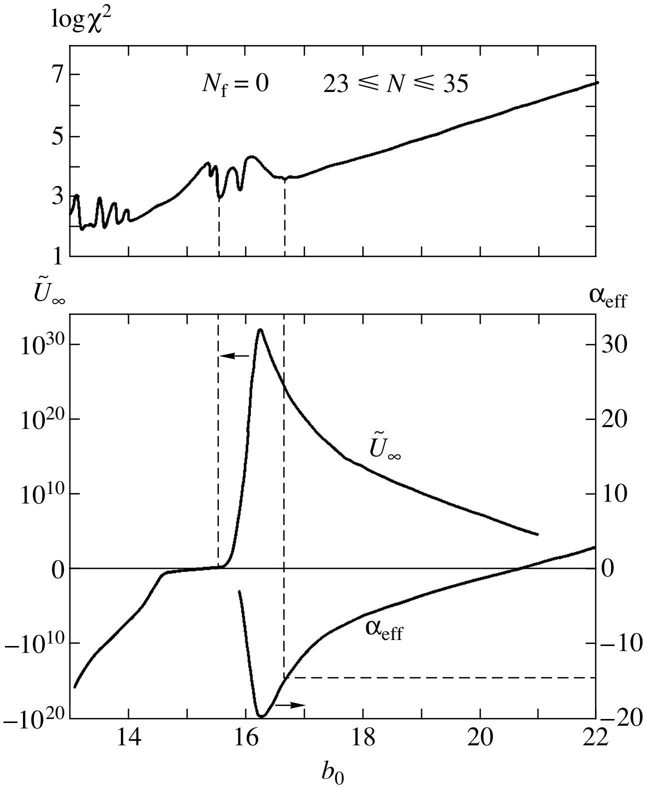


Fig. 21. Curves of χ^2, α_{eff}, and $\tilde{U}_\infty$ versus b_0 obtained for QCD. The minima at $b_0 = 15.4$ and 15.9 are interpreted as satellites moving with the main minimum at $b_0 = 15.5$.

---

[20] The parameter c was estimated as the product of the squared 't Hooft constant c_H in one-instanton contribution (4.5.6) ($c_H^2 \sim 10^{-5}$ and 10^{-4} for $N_f = 0$ and 3, respectively) with the dimensionless integral of the instanton configuration. The latter factor is relatively large (its characteristic value is $16\pi^2$).

the right-hand minimum of χ^2. The values of α estimated by these methods agree only for $\tilde{N}$ close to the optimal value $\tilde{N} = 1.58$ (Fig. 21) and tend to disagree as the difference between $\tilde{N}$ and this value increases.

The resulting value of α cannot be accepted as final. First, a large value of α may be indicative of exponential behavior. Second, since $\Gamma(\alpha)$ has poles at $\alpha = 0, -1, -2, \ldots$ (see (6.12)), the leading contribution to the asymptotic behavior of U_N may vanish, and the result may correspond, for example, to α' in (6.15). For this reason, the function $W(g) = g^{n_s}\beta(g)$ is introduced, and the integer parameter n_s is increased until the exponent $\alpha_W = \alpha + n_s$ becomes positive. The results obtained by this method (Fig. 22a) demonstrate that the true behavior a power law with a large noninteger negative exponent rather than an exponential. (If $\alpha = -n$, the exponent would behave as illustrated by the inset.) Each point in Fig. 22a is obtained by independent optimization in $\tilde{N}$. The optimal $\tilde{N}$ decreases monotonically with increasing n_s. The uncertainty of the results is primarily due to their dependence on the lower limit of the averaging interval $N_{\min} \le N \le N_{\max}$. The higher lying data points in Fig. 22a correspond to small $N_{\min}$ and minimum values of χ^2 on the order of 10^6. As $N_{\min}$ increases, α decreases monotonically until χ^2 reaches a value on the order of 10^3 (lower lying data points). With a further increase in $N_{\min}$, the pattern of χ^2 minima becomes indistinct and the uncertainty of the results sharply increases. The value of α is then allowed to decrease further until $\chi^2 \sim 10$ is reached as required, and this is taken into account in error estimation. Even though the uncertainty in $\bar{\beta}_\infty$ amounts to several orders of magnitude (Fig. 22b), the most probable value must be on the order of 10^5 to be consistent with most data. Thus,

$$\alpha = -13 \pm 2, \quad \bar{\beta}_\infty \sim 10^5. \tag{8.18}$$

for $N_f = 0$. For $N_f = 3$, the result is $\alpha = -12 \pm 3$, and the same most probable value is obtained (while $\bar{\beta}_\infty$ is scattered between 1 and 10^7). The robustness results under the change in summation procedure means that their uncertainty has been adequately estimated.

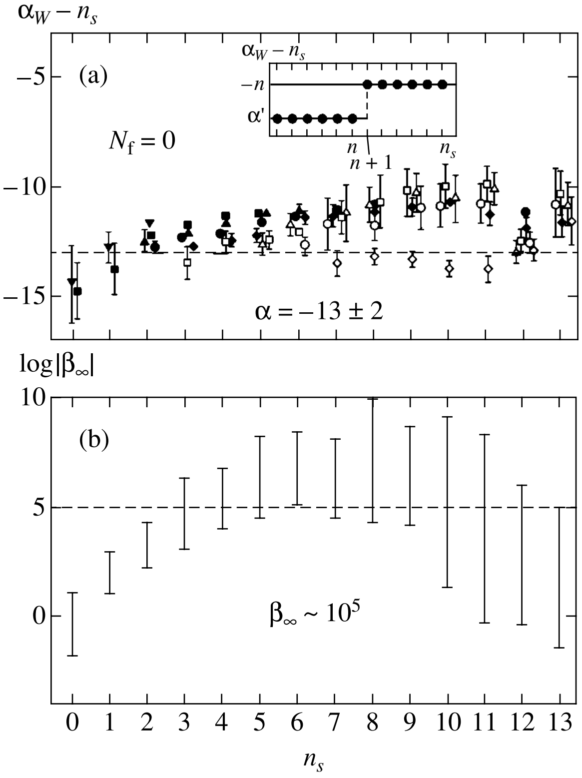


Fig. 22. (a) Exponent α_W for QCD obtained by summation of series for $W(g) = g^{n_s}\beta(g)$ versus n_s for different intervals $N_{\min} \le N \le N_{\max}$: ▼ corresponds to $N_{\min} = 22 + n_s$, $N_{\max} = 35 + n_s$; ■, ▲, ●, □, △, ○, and ◇ correspond to $N_{\min}$ increased in unit steps. (b) $\bar{\beta}_\infty$ versus n_s.

While the uncertainty in $\bar{\beta}_\infty$ is large, the corresponding uncertainty in the β function to relatively small: the one-loop law $\beta_2 g^2$ is matched with asymptotic expression (8.17) at $g^* \sim 2$, and $\bar{\beta}_\infty$ changes by four orders of magnitude as g^* changes by a factor of two. When α_W is negative, the sign of $\bar{\beta}_\infty$ is indeterminate, because the error in α is large and the factor $\Gamma(\alpha)$ in Eq. (6.12) is alternating, but this sign is definitely negative for positive α_W (large n_s). Figure 15c illustrates the behavior of β-function at $g < 0$ (solid curve). Its analytic continuation to positive g is expected to exhibit qualitatively similar behavior, but the sign of (8.17) may change (dashed curve).[21] Nevertheless, the effective coupling constant exhibits consistent behavior as a function of the length scale L (Fig. 15d). In the one-loop approximation, $g(L)$ has a pole at $L = L_0 = 1/\Lambda_{QCD}$ (dashed curve). For the obtained β function, $g(L)$ increases in the neighborhood of L_0 until a value on the order of g^* is reached (see Fig. 15c) and then either becomes constant (if $\beta_\infty > 0$) or nearly constant, increasing as $(\ln L)^{0.07}$ (if $\beta_\infty < 0$).

In the weak-coupling region, the quark–quark interaction potential $V(L)$ is described by the modified Coulomb law $\bar{g}^2(L)/L$, and the sharp increase in $\bar{g}(L)$ in the neighborhood of $L = L_0$ points to a tendency to confinement. In the strong-coupling region, the relation between $V(L)$ and $\bar{g}(L)$ is not known. However, an

---

[21] In particular, $\beta_\infty = \bar{\beta}_\infty \cos\pi\alpha$ if the Borel integral is interpreted in the sense of the Cauchy principal value.

analogous result was obtained by Wilson [138] in lattice QCD:

$$V(L) = \frac{\ln 3\bar{g}^2(a)}{a^2}L, \quad \bar{g}(a) \gg 1, \qquad (8.19)$$

where a is the lattice constant. Since the result must be independent of a, the β function in the strong-coupling region may be estimated as $\beta(g) \sim g\ln g$ [139], which is, however, incorrect. The transverse size of the string estimated for $a \gg 1/\Lambda_{QCD}$ is on the order of a, which is much larger than its actual physical size ($\sim 1/\Lambda_{QCD}$). This means that lattice effects are so strong that there is no reason to expect that the result will be independent of a. With regard to $a \ll 1/\Lambda_{QCD}$, there is some reason to these expectations, but Eq. (8.19) does not apply since the coupling constant $\bar{g}(a)$ is small. Thus, Eq. (8.19) may be valid only for $a \sim 1/\Lambda_{QCD}$. In the plateau region, $\bar{g}(L) \sim \sqrt{2 \cdot 16\pi^2} \sim 20$, and the sharp increase in $\bar{g}(L)$ in the neighborhood of $L = L_0$ (Fig. 15d) implies that the conditions $a \sim 1/\Lambda_{QCD}$ and $\bar{g}(a) \gg 1$ are compatible; i.e., lattice formula (8.19) can be applied to actual QCD.

9. HIGH-ORDER CORRECTIONS TO THE LIPATOV ASYMPTOTICS

As noted above, corrections to Lipatov asymptotic form (2.5) can be represented by a regular expansion in terms of $1/N$:

$$W_N = ca^N\Gamma(N+b) \times \left\{1 + \frac{A_1}{N} + \frac{A_2}{N^2} + \ldots + \frac{A_K}{N^K} + \ldots\right\}. \qquad (9.1)$$

Knowledge of all coefficients A_K is equivalent to knowledge of the exact coefficient function W_N, and their calculation offers an alternative to direct calculation of low-order diagrams [81, 106, 125, 133, 137]. Currently, the lowest order corrections are known in φ^4 theory [126] and a number of quantum-mechanical problems [6, 140].

It was shown in [141] that series (9.1) is factorially divergent, and high-order expansion coefficients can be calculated by using a procedure analogous to Lipatov's method: an exact expression for the Kth coefficient can be written as a functional integral and found by the saddle-point method for large K. Typically, A_K has the asymptotic form

$$A_K = \tilde{c}\left(\ln\frac{S_1}{S_0}\right)^{-K}\Gamma\left(K + \frac{r' - r}{2}\right), \qquad (9.2)$$

where S_0 and S_1 are the values of action for the first and second instantons in the field theory under analysis, and r and r' denote the corresponding number of zero modes. The instantons are enumerated in the order of increasing action.

Detailed calculations of the asymptotic form of A_K for the n-component φ^4 theory were presented in [33]. Available information about higher order instantons in φ^4 theory is incomplete. However, it is reasonable to assume that the role of the second instanton is played by a combination of two elementary instantons [33, 142]. Then, expression (9.2) must be modified, because it is correct only when the equidistribution principle is valid (see Section 4.1), i.e., when the fluctuation modes can be strictly classified as zero and oscillatory ones. For two-instanton configurations, there must exist a soft mode that corresponds to variation of the distance between the elementary instantons and can be reduced to oscillation in a potential well with nonanalytic minimum. Accordingly, logarithmic and power-law corrections appear in (9.2) if $d = 1, 2, 3$ and $d = 4$, respectively.

If $d = 1$, then the asymptotic form of the coefficients A_K corresponding to the M-point Green function $G_M(g)$ is

$$A_K = -\frac{2^{-M/2}}{(\pi/2)\Gamma(n/2)}\left(\frac{3}{2\ln 2}\right)^{n/2} \times\Gamma\left(K + \frac{n}{2}\right)(\ln 2)^{-K}[\ln K + C], \qquad (9.3)$$

$$C = C_{\mathrm{E}} + \ln\left(\frac{6}{\ln 2}\right) + \frac{\psi(1/2) - \psi(n/2)}{2},$$

where C_{E} is Euler's constant and $\psi(x)$ is the logarithmic derivative of the gamma function. If $d = 2$, then

$$A_K = -\frac{2^{-M/2}}{19.7}\frac{(0.702)^n}{\Gamma(n/2)} \times\Gamma\left(K + \frac{n+1}{2}\right)(\ln 2)^{-K}\ln^2 K \qquad (9.4)$$

to logarithmic accuracy. Similarly,

$$A_K = -\frac{2^{-M/2}}{2.12}\frac{(0.704)^n}{\Gamma(n/2)} \times\Gamma\left(K + \frac{n+2}{2}\right)(\ln 2)^{-K}\ln^3 K \qquad (9.5)$$

for $d = 3$. The results obtained for $d = 4$ depend on the coordinates contained in the Green functions and have cumbersome expressions [33]. They can be simplified by changing to the representation in terms of momenta p_i corresponding to the symmetric point ($p_i \sim p$):

$$A_K = Be^{\nu\ln(\mu/p)}\Gamma\left(K + \frac{n+4}{2} + \nu\right)(\ln 2)^{-K}, \qquad (9.6)$$

where μ is a constant determined by the charge normalization condition, $\nu = (n+8)/3$, and the values of B are listed in Table 3. In the scalar theory ($n = 1$), the leading contribution to the asymptotic expression vanishes, and the asymptotic behavior is expected to be determined by the next-order term in $1/K$:

$$A_K = \mathrm{const}\, e^{\nu\ln(\mu/p)} \times\Gamma\left(K+\frac{n+4}{2}+\nu-1\right)(\ln 2)^{-K}. \tag{9.7}$$

The results for the logarithm of the vacuum integral $Z_0(g)$ are formally derived by setting $M = 0$ and introducing a factor of 1/2 in (9.3)–(9.6). In particular, the following result is obtained for the ground-state energy of the anharmonic oscillator ($d = 1$, $n = 1$):

$$A_K = -\frac{\ln K + 2.74}{3.78}\Gamma\left(K+\frac{1}{2}\right)(\ln 2)^{-K}. \tag{9.8}$$

Figure 23 compares this prediction with numerical results obtained in [6].

If $d = 1$, then the entire instanton spectrum can be represented by combinations of elementary instantons. If $d \geq 2$, then there may exist a nonspherically symmetric instanton with action lower than $2S_0$. In this case, there are no soft modes, and formula (9.2) with $r' - r = d(d-1)/2$ is valid, because an asymmetric instanton is associated with $d(d-1)/2$ additional modes corresponding to rotations in the coordinate space. Since modes of this kind have never been considered, the calculation of the constant $\tilde{c}$ in (9.2) is a technically nontrivial problem. The technique of integration over these modes developed in [33] must be instrumental in quantum electrodynamics, where even the first instanton is asymmetric [23].

10. OUTLOOK

Finally, let us discuss the most promising lines of further research.

10.1. Calculation of c in the Lipatov Asymptotics

Complete Lipatov asymptotic forms are known only in φ^4 theory and a number of quantum-mechanical problems. In other models, the common factor c has yet to be calculated. For the Gell-Mann–Low function in QCD, the factor c has been calculated only in the case of $SU(2)$ symmetry [17]. Note that the calculation made use of an unconventional definition of the β function. Therefore, consistency of the asymptotic form of β_N with the renormalization scheme used in actual diagrammatic calculations remains an open question. The factor c has been formally calculated for the quark–quark correlation function in QCD [24]. However, since the procedure used to eliminate divergences in (4.5.13) evokes doubts [48], even the general form of the result may be revised.

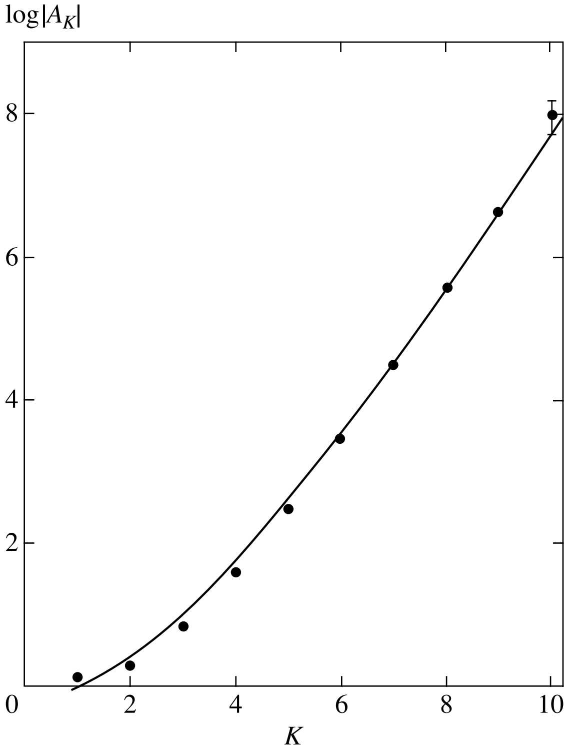


Fig. 23. Predictions of asymptotic formula (9.8) (curve) compared with coefficients A_K calculated numerically in [6] (symbols).

10.2. A Priori Proofs of Absence of Renormalons

A proof of this kind has been found only for φ^4 theory. The constructive scheme proposed in Section 5.3 for eliminating renormalon singularities in QED and QCD is substantiated by results presented in Section 8. However, it relies on approximate determination of the Gell-Mann–Low function, which may seem questionable to a skeptical reader. Therefore, extension of the analysis presented in [48] to other field theories is highly desirable.

Table 3. Parameter B in formula (9.6)

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

10.3. Development and Application of Highly Accurate Summation Methods

In the scheme of conformal–Borel technique presented in Section 6.1, the cut in the Borel plane extends from $-\infty$ to the nearest instanton singularity S_0. However, the cut can be extended to point S^* such that $S_0 < S^* < 0$. In this case, all singularities of the Borel transform remain on the boundary of a unit circle in the u plane, and the resummed series is convergent at every point in the Borel plane that does not lie on the cut. If $S_0 < S^* < 0$, then the results are independent of S^*. This empirical observation made in [115] implies that knowledge of the exact value of S_0 is not required, and no quantitative information about the Lipatov asymptotic form is used in the scheme proposed by Le Guillou and Zinn-Justin [35]. I believe that interpolation of the coefficient function and explicit use of the asymptotic behavior at strong coupling will substantially improve the accuracy of calculations of critical exponents based on available information. This example illustrates the inefficiency of the current use of information that requires enormous labor resources to be acquired.

Additional improvement of efficiency can be achieved by using information concerning high-order corrections to the Lipatov asymptotic form. The scheme described in Section 9 facilitates the calculation of several parameters characterizing the coefficient function. In terms of efficiency, this is equivalent to advancing by several orders in perturbation theory, whereas advancement to the next order in diagrammatic calculations requires about ten years.

The method for finding strong-coupling asymptotics described in Section 6.2 is effective when information is scarce, but cannot be classified as a highly accurate one. When more information is available, construction of Padé approximants for the coefficient function [13] is a more effective tool. Preliminary studies show that this method can be combined with some strategy for selecting most suitable Padé approximants.

10.4. Summation of Nonalternating Series

In essence, the analysis presented in Section 7 solves the problem of non-Borel-summability for the most interesting problems. However, the summation schemes formulated therein are insufficiently accurate, and improved methods should be developed.

The summation of QCD perturbation series is performed in Section 8.3 without invoking the technique developed in Section 7. A certain trick is used to circumvent the problem, and the resulting exponent α is correct, whereas β_∞ is determined only up to an order of magnitude. Currently, this rough approximation is acceptable in view of large uncertainty in β_∞ (see Fig. 22b). However, it is strongly recommended that a test summation be performed “by following all the rules” and essential uncertainties be analyzed.

In the confinement problem, summation of expansions for anomalous dimensions is required. The formation of a string -like “flux tube” between quarks is not controlled by the β function, being determined by properties of correlation functions, which depend on the values of anomalous dimensions in the “plateau” region of the coupling constant (see Fig 15d).

10.5. Analytical Methods for Strong-Coupling Problems

The exponent α is close to unity in both QCD and φ^4 theory. Moreover, there are reasons to believe that its exact value is $\alpha = 1$. Simple results of this kind must be obtainable by analytical methods. Since a known result is always easier to substantiate than to obtain, there are grounds for an optimistic outlook. Once the equality $\alpha = 1$ is proved, the accuracy of analysis of strong-coupling asymptotics will substantially improve, because the number of parameters to be determined will reduce from two to one. It is obvious that progress in this area will be stimulated by acquiring additional “experimental” information concerning strong-coupling asymptotics.

10.6. Applications to the Theory of Disordered Systems

The theory of disordered systems is unique in that high-order contributions are essential even in the weak-coupling region. Description of a particle moving in a Gaussian random field can be rigorously reformulated as a φ^4 theory with “incorrect” sign of the coupling constant [103, 121–123]. In formally unstable field theories of this kind, nonperturbative contributions of the form $\exp(-a/g)$ play an important role and can be found by summing perturbation series. One example is the fluctuation tail of density of states [143], which is directly related to the Lipatov asymptotic form [8, 144]. Combination of instanton calculations with parquet approximation can be used to develop a complete theory of density of states for a disordered system in the $(4 - \epsilon)$-dimensional space [8]. Next in order is the development of an analogous kinetic theory for calculating transport disordered systems, which requires an analysis of a φ^4-type theory with two vector fields [122, 123]. The qualitative importance of high-order perturbative contributions for such an analysis is due to the purely nonperturbative nature of the diffusion pole in the localized phase [145, 146]. If all characteristics of the pole can be elucidated in the framework of the instanton method, then an explanation will be found for the “simple” critical exponents obtained in the symmetry-based approach to the theory of the Anderson transition [147].

ACKNOWLEDGMENTS

I thank M.V. Sadovskii and A.I. Sokolov, who read a preliminary version of the manuscript and made important remarks; L.N. Lipatov for stimulating discussions; K.B. Varnashev for help in selecting referenced papers; and participants of seminars at the Kapitza Institute for Physical Problems, Lebedev Physical Institute, Institute of Theoretical and Experimental Physics, and Konstantinov Institute of Nuclear Physics for interest in this study and numerous discussions.

This work was supported by the Russian Foundation for Basic Research, project no. 03-02-17519.

REFERENCES

1. J.-P. Ramis, *Series divergentes et theories asymptotiques* (Am. Math. Soc., Providence, RI, 1984; Inst. Komp. Issled., Izhevsk, 2002).
2. N. N. Bogolyubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields*, 3rd ed. (Nauka, Moscow, 1976; Wiley, New York, 1980).
3. V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics*, 2nd ed. (Nauka, Moscow, 1980; Pergamon, Oxford, 1982).
4. A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskiĭ, *Methods of Quantum Field Theory in Statistical Physics* (Fizmatgiz, Moscow, 1962; Prentice-Hall, Englewood Cliffs, N.J., 1963).
5. F. J. Dyson, Phys. Rev. **85**, 631 (1952).
6. C. M. Bender and T. T. Wu, Phys. Rev. **184**, 1231 (1969); Phys. Rev. D **7**, 1620 (1973).
7. L. N. Lipatov, Zh. Éksp. Teor. Fiz. **72**, 411 (1977) [Sov. Phys. JETP **45**, 216 (1977)].
8. I. M. Suslov, Usp. Fiz. Nauk **168**, 503 (1998) [Phys. Usp. **41**, 441 (1998)].
9. L. D. Landau, A. A. Abrikosov, and I. M. Khalatnikov, Dokl. Akad. Nauk SSSR **95**, 497, 773, 1177 (1954).
10. N. N. Bogolyubov and D. V. Shirkov, *Quantum Fields* (Nauka, Moscow, 1993) [in Russian].
11. V. S. Popov, V. L. Eletskiĭ, and A. V. Turbiner, Zh. Éksp. Teor. Fiz. **74**, 445 (1978) [Sov. Phys. JETP **47**, 232 (1978)].
12. D. I. Kazakov, O. V. Tarasov, and D. V. Shirkov, Teor. Mat. Fiz. **38**, 15 (1979).
13. Yu. A. Kubyshin, Teor. Mat. Fiz. **58**, 137 (1984).
14. E. Brezin, J. C. Le Guillou, and J. Zinn-Justin, Phys. Rev. D **15**, 1544 (1977).
15. G. Parisi, Phys. Lett. B **66B**, 382 (1977).
16. A. P. Bukhvostov and L. N. Lipatov, Zh. Éksp. Teor. Fiz. **73**, 1658 (1977) [Sov. Phys. JETP **46**, 871 (1977)]; C. Itzykson, G. Parisi, and J. B. Zuber, Phys. Rev. Lett. **38**, 306 (1977).
17. E. B. Bogomolny and V. A. Fateyev, Phys. Lett. B **71B**, 93 (1977).
18. L. N. Lipatov, A. P. Bukhvostov, and E. I. Malkov, Phys. Rev. D **19**, 2974 (1979).
19. *Large Order Behavior of Perturbation Theory*, Ed. by J. C. Le Guillou and J. Zinn-Justin (North-Holland, Amsterdam, 1990).
20. E. B. Bogomolny, V. A. Fateyev, and L. M. Lipatov, in *Soviet Science Reviews: Physics*, Ed. by I. M. Khalatnikov (Harwood Academic, New York, 1980), Vol. 2, p. 247.
21. J. Zinn-Justin, Phys. Rep. **70**, 109 (1981).
22. C. Itzykson, G. Parisi, and J. B. Zuber, Phys. Rev. D **16**, 996 (1977); R. Balian, C. Itzykson, G. Parisi, and J. B. Zuber, Phys. Rev. D **17**, 1041 (1978).
23. E. B. Bogomolny and V. A. Fateyev, Phys. Lett. B **76B**, 210 (1978).
24. I. I. Balitsky, Phys. Lett. B **273**, 282 (1991).
25. P. G. Silvestrov, Phys. Rev. D **51**, 6587 (1995).
26. S. V. Faleev and P. G. Silvestrov, Nucl. Phys. B **463**, 489 (1996).
27. G. H. Hardy, *Divergent Series*, 2nd ed. (Clarendon, Oxford, 1956; Inostrannaya Literatura, Moscow, 1951).
28. L. D. Kudryavtsev, *Mathematical Analysis* (Vysshaya Shkola, Moscow, 1973), Vol. 1 [in Russian].
29. E. Borel, *Memoire sur les series divergentes* (Gauthier-Villars, Paris, 1899).
30. H. Poincaré, Acta Math. **5**, 240 (1884).
31. E. B. Bogomolny, Phys. Lett. B **67B**, 193 (1977).
32. G. Parisi, Phys. Lett. B **66B**, 167 (1977).
33. D. A. Lobaskin and I. M. Suslov, Zh. Éksp. Teor. Fiz. **126**, 268 (2004) [JETP **99**, 234 (2004)].
34. E. Brezin and G. Parisi, J. Stat. Phys. **19**, 269 (1978).
35. J. C. Le Guillou and J. Zinn-Justin, Phys. Rev. Lett. **39**, 95 (1977); Phys. Rev. B **21**, 3976 (1980); R. Guida and J. Zinn-Justin, J. Phys. A **31**, 8103 (1998).
36. I. M. Suslov, Zh. Éksp. Teor. Fiz. **106**, 560 (1994) [JETP **79**, 307 (1994)].
37. I. M. Suslov, Zh. Éksp. Teor. Fiz. **111**, 220 (1997) [JETP **84**, 120 (1997)].
38. I. M. Suslov, Zh. Éksp. Teor. Fiz. **111**, 1896 (1997) [JETP **84**, 1036 (1997)].
39. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics*, 3rd ed. (Nauka, Moscow, 1976; Pergamon, Oxford, 1980).
40. F. A. Berezin, *Method of Second Quantization* (Nauka, Moscow, 1965; Academic, New York, 1966).
41. R. Rajaraman, *Solitons and Instantons: An Introduction to Solitons and Instantons in Quantum Field Theory* (North-Holland, Amsterdam, 1982; Mir, Moscow, 1985).
42. I. M. Suslov, Pis'ma Zh. Éksp. Teor. Fiz. **74**, 211 (2001) [JETP Lett. **74**, 191 (2001)].
43. E. Brezin, G. Parisi, and J. Zinn-Justin, Phys. Rev. D **16**, 408 (1977).
44. A. A. Belavin, A. M. Polyakov, A. S. Shwartz, and Yu. S. Tyupkin, Phys. Lett. B **59B**, 85 (1975).
45. G. 't Hooft, Phys. Rev. D **14**, 3432 (1976).
46. C. Bernard, Phys. Rev. D **19**, 3013 (1979).
47. I. M. Suslov, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 387 (2002) [JETP Lett. **76**, 327 (2002)].
48. I. M. Suslov, Zh. Éksp. Teor. Fiz. **116**, 369 (1999) [JETP **89**, 197 (1999)].

49. B. Lautrup, Phys. Lett. B **69B**, 109 (1977).
50. G. 't Hooft, in *The Whys of Subnuclear Physics: Proceedings of the 1977 International School of Subnuclear Physics, Erice, Trapani, Sicily, 1977*, Ed. by A. Zichichi (Plenum, New York, 1979).
51. Yu. V. Sidorov, M. V. Fedoryuk, and M. I. Shabunin, in *Lectures on the Theory of Functions of a Complex Variable* (Nauka, Moscow, 1976) [in Russian].
52. S. V. Faleev and P. G. Silvestrov, Nucl. Phys. B **507**, 379 (1997).
53. M. Beneke, Phys. Rep. **317**, 1 (1999).
54. G. Parisi, Phys. Lett. B **76B**, 65 (1978); Nucl. Phys. B **150**, 163 (1979).
55. G. Parisi, Phys. Rep. **49**, 215 (1979).
56. F. David, Nucl. Phys. B **209**, 433 (1982); **234**, 237 (1984); **263**, 637 (1986).
57. M. C. Bergere and F. David, Phys. Lett. B **135B**, 412 (1984).
58. A. H. Mueller, Nucl. Phys. B **250**, 327 (1985).
59. V. I. Zakharov, Nucl. Phys. B **385**, 452 (1992).
60. M. Beneke *et al.*, Phys. Lett. B **307**, 154 (1993); **348**, 613 (1995); Nucl. Phys. B **452**, 563 (1995); **472**, 529 (1996); Phys. Rev. D **52**, 3929 (1995).
61. G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers*, 2nd ed. (McGraw-Hill, New York, 1968; Nauka, Moscow, 1977).
62. M. A. Evgrafov, *Analytic Functions*, 2nd ed. (Nauka, Moscow, 1968; Saunders, Philadelphia, 1966).
63. I. M. Suslov, Zh. Éksp. Teor. Fiz. **126**, 542 (2004) [JETP **99**, 474 (2004)].
64. I. M. Suslov, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 315 (2000) [JETP Lett. **71**, 217 (2000)].
65. I. M. Suslov, Zh. Éksp. Teor. Fiz. **120**, 5 (2001) [JETP **93**, 1 (2001)].
66. G. A. Baker, Jr., B. G. Nickel, M. S. Green, and D. I. Meiron, Phys. Rev. Lett. **36**, 1351 (1976); Phys. Rev. B **17**, 1365 (1978).
67. J. S. R. Chisholm, Math. Comput. **27**, 841 (1973).
68. G. A. Baker, Jr. and P. Graves-Morris, *Pade Approximants* (Addison-Wesley, Reading, MA, 1981).
69. I. O. Mayer, J. Phys. A **22**, 2815 (1989).
70. S. A. Antonenko and A. I. Sokolov, Phys. Rev. B **49**, 15901 (1994).
71. I. O. Mayer, Teor. Mat. Fiz. **75**, 234 (1988).
72. R. P. Feynman and H. Kleinert, Phys. Rev. A **34**, 5080 (1986); H. Kleinert, Phys. Lett. A **173**, 332 (1993).
73. H. Kleinert and V. Schulte-Frohlinde, *Critical Properties of* φ^4*-Theories* (World Sci., Singapore, 2001).
74. H. Kleinert, Phys. Rev. Lett. **75**, 2787 (1995); Phys. Rev. D **57**, 2264 (1998); Phys. Lett. B **434**, 74 (1998).
75. H. Kleinert, Phys. Rev. D **60**, 085001 (1999).
76. E. Brezin, J. C. Le Guillou, and J. Zinn-Justin, in *Phase Transitions and Critical Phenomena*, Ed. by C. Domb and M. S. Green (Academic, New York, 1976), Vol. 6.
77. J. Zinn-Justin, *Quantum Field Theory and Critical Phenomena* (Clarendon, Oxford, 2002).
78. S. A. Antonenko and A. I. Sokolov, Phys. Rev. E **51**, 1894 (1995).
79. A. I. Sokolov, Fiz. Tverd. Tela (St. Petersburg) **40**, 1284 (1998) [Phys. Solid State **40**, 1169 (1998)]; A. I. Sokolov, E. V. Orlov, V. A. Ul'kov, and S. S. Kashtanov, Phys. Rev. E **60**, 1344 (1999).
80. R. Guida and J. Zinn-Justin, J. Phys. A **31**, 8103 (1998).
81. E. V. Orlov and A. I. Sokolov, Fiz. Tverd. Tela (St. Petersburg) **42**, 2087 (2000) [Phys. Solid State **42**, 2151 (2000)].
82. P. Calabrese, E. V. Orlov, D. V. Pakhnin, and A. I. Sokolov, Phys. Rev. B **70**, 094425 (2004).
83. G. Grinstein and A. Luther, Phys. Rev. B **13**, 1329 (1976); A. Aharony, Phys. Rev. B **13**, 2092 (1976).
84. I. O. Mayer, A. I. Sokolov, and B. N. Shalaev, Ferroelectrics **95**, 93 (1989).
85. D. V. Pakhnin and A. I. Sokolov, Phys. Rev. B **61**, 15130 (2000).
86. J. M. Carmona, A. Pelissetto, and E. Vicari, Phys. Rev. B **61**, 15136 (2000).
87. A. Pelissetto and E. Vicari, Phys. Rev. B **62**, 6393 (2000).
88. D. V. Pakhnin, A. I. Sokolov, and B. N. Shalaev, Pis'ma Zh. Éksp. Teor. Fiz. **75**, 459 (2002) [JETP Lett. **75**, 387 (2002)].
89. Yu. Holovatch, V. Blavats'ka, M. Dudka, *et al.*, Int. J. Mod. Phys. B **16**, 4027 (2002).
90. A. Pelissetto, P. Rossi, and E. Vicari, Phys. Rev. B **63**, 140414 (2001); **65**, 020403 (2002).
91. P. Calabrese, P. Parruccini, and A. I. Sokolov, Phys. Rev. B **66**, 180403 (2002); **68**, 094415 (2003).
92. R. Folk, Yu. Holovatch, and T. Yavors'kii, Phys. Rev. B **62**, 12195 (2000).
93. M. Dudka, Yu. Holovatch, and T. Yavors'kii, J. Phys. A **37**, 10727 (2004).
94. A. I. Sokolov and K. B. Varnashev, Phys. Rev. B **59**, 8363 (1999).
95. A. I. Mudrov and K. B. Varnashev, Phys. Rev. B **57**, 3562 (1998); **57**, 5704 (1998); **64**, 214423 (2001).
96. A. Pelissetto and E. Vicari, Phys. Rep. **368**, 549 (2002).
97. P. Calabrese, E. V. Orlov, P. Parruccini, and A. I. Sokolov, Phys. Rev. B **67**, 024413 (2003).
98. V. V. Prudnikov, S. V. Belim, A. V. Ivanov, *et al.*, Zh. Éksp. Teor. Fiz. **114**, 972 (1998) [JETP **87**, 527 (1998)].
99. V. V. Prudnikov, S. V. Belim, E. V. Osintsev, and A. A. Fedorenko, Fiz. Tverd. Tela (St. Petersburg) **40**, 1526 (1998) [Phys. Solid State **40**, 1383 (1998)].
100. V. V. Prudnikov, P. V. Prudnikov, and A. A. Fedorenko, J. Phys. A **32**, 8587 (1999); S. V. Belim, Pis'ma Zh. Éksp. Teor. Fiz. **77**, 118 (2003) [JETP Lett. **77**, 112 (2003)]; Pis'ma Zh. Éksp. Teor. Fiz. **77**, 509 (2003) [JETP Lett. **77**, 434 (2003)].
101. V. V. Prudnikov, P. V. Prudnikov, and A. A. Fedorenko, Pis'ma Zh. Éksp. Teor. Fiz. **73**, 153 (2001) [JETP Lett. **73**, 135 (2001)]; P. V. Prudnikov and V. V. Prudnikov, Zh. Éksp. Teor. Fiz. **122**, 636 (2002) [JETP **95**, 550 (2002)].

102. K. G. Wilson and J. Kogut, Phys. Rep. **12C**, 75 (1975); *The Renormalization Group and the ϵ-Expansion* (Mir, Moscow, 1975).

103. S. Ma, *Modern Theory of Critical Phenomena* (Benjamin, Reading, Mass., 1976; Mir, Moscow, 1980).

104. A. A. Vladimirov, D. I. Kazakov, and O. V. Tarasov, Zh. Éksp. Teor. Fiz. **77**, 1035 (1979) [Sov. Phys. JETP **50**, 521 (1979)].

105. J. C. Le Guillou and J. Zinn-Justin, J. Phys. Lett. **46**, L131 (1985); J. Phys. (Paris) **48**, 19 (1987); **50**, 1365 (1989).

106. H. Kleinert, J. Neu, V. Schulte-Frohlinde, *et al.*, Phys. Lett. B **272**, 39 (1991); **319**, 545 (1993).

107. H. Kleinert and V. Schulte-Frohlinde, Phys. Lett. B **342**, 284 (1995).

108. B. N. Shalaev, S. A. Antonenko, and A. I. Sokolov, Phys. Lett. A **230**, 105 (1997).

109. H. Kleinert, S. Thoms, and V. Schulte-Frohlinde, Phys. Rev. B **56**, 14428 (1997).

110. R. Folk, Yu. Holovatch, and T. Yavors'kii, Phys. Rev. B **61**, 15114 (2000).

111. R. Folk, Y. Holovatch, and T. Yavorskiĭ, Usp. Fiz. Nauk **173**, 175 (2003) [Phys. Usp. **46**, 169 (2003)].

112. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes* (Cambridge Univ. Press, Cambridge, 1988).

113. I. M. Suslov, Zh. Éksp. Teor. Fiz. **122**, 696 (2002) [JETP **95**, 601 (2002)].

114. D. I. Kazakov and V. S. Popov, Zh. Éksp. Teor. Fiz. **122**, 675 (2002) [JETP **95**, 581 (2002)].

115. A. I. Mudrov and K. B. Varnashev, Phys. Rev. E **58**, 5371 (1998).

116. U. D. Jentschura and J. Zinn-Justin, J. Phys. A **34**, L253 (2001).

117. D. I. Kazakov and V. S. Popov, Pis'ma Zh. Éksp. Teor. Fiz. **77**, 547 (2003) [JETP Lett. **77**, 453 (2003)].

118. J. Zinn-Justin, *Quantum Field Theory and Critical Phenomena* (Clarendon, Oxford, 1989), Chap. 40.

119. W. Koppenfels and F. Stallmann, *Praxis der Konformen Abbildung* (Springer, Berlin, 1959; Inostrannaya Literatura, Moscow, 1963).

120. A. N. Sissakian, I. L. Solovtsov, and O. P. Solovtsova, Phys. Lett. B **321**, 381 (1994).

121. D. J. Thouless, J. Phys. C **8**, 1803 (1975).

122. A. Nitzan, K. F. Freed, and M. N. Cohen, Phys. Rev. B **15**, 4476 (1977).

123. M. V. Sadovskii, Usp. Fiz. Nauk **133**, 223 (1981) [Sov. Phys. Usp. **24**, 96 (1981)]; Zh. Éksp. Teor. Fiz. **70**, 1936 (1976) [Sov. Phys. JETP **43**, 1008 (1976)].

124. A. A. Vladimirov and D. V. Shirkov, Usp. Fiz. Nauk **129**, 407 (1979) [Sov. Phys. Usp. **22**, 860 (1979)].

125. F. M. Dittes, Yu. A. Kubyshin, and O. V. Tarasov, Teor. Mat. Fiz. **37**, 66 (1978).

126. Yu. A. Kubyshin, Teor. Mat. Fiz. **57**, 363 (1983).

127. L. N. Lipatov, Zh. Éksp. Teor. Fiz. **71**, 2010 (1976) [Sov. Phys. JETP **44**, 1055 (1976)].

128. J. Frolich, Nucl. Phys. B **200**, 281 (1982).

129. J. P. Eckmann and R. Epstein, Comm. Math. Soc. **64**, 95 (1979).

130. M. Aizenman, Comm. Math. Soc. **86**, 1 (1982).

131. M. Consoli and P. M. Stevenson, Z. Phys. C **63**, 427 (1994).

132. A. Agodi, G. Andronico, P. Cea, *et al.*, Mod. Phys. Lett. A **12**, 1011 (1997).

133. S. G. Gorishny, A. L. Kataev, S. A. Larin, and L. R. Surguladze, Phys. Lett. B **256**, 81 (1991).

134. N. V. Krasnikov, Nucl. Phys. B **192**, 497 (1981); H. Yamagishi, Phys. Rev. D **25**, 464 (1982).

135. S. Kim, J. B. Kogut, and M. P. Lombardo, Phys. Lett. B **502**, 345 (2001).

136. V. Azcoiti, Nucl. Phys., Proc. Suppl. **53**, 148 (1997).

137. T. van Ritbergen, J. A. M. Vermaseren, and S. A. Larin, Phys. Lett. B **400**, 379 (1997).

138. K. Wilson, Phys. Rev. D **10**, 2445 (1974).

139. C. Callan, R. Dashen, and D. Gross, Phys. Rev. D **20**, 3279 (1979).

140. S. V. Faleev and P. G. Silvestrov, Phys. Lett. A **197**, 372 (1995).

141. I. M. Suslov, Zh. Éksp. Teor. Fiz. **117**, 659 (2000) [JETP **90**, 571 (2000)].

142. 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); G. L. Alfimov, Izv. Ross. Akad. Nauk, Ser. Fiz. **60**, 12 (1996).

143. I. M. Lifshits, Usp. Fiz. Nauk **83**, 617 (1964) [Sov. Phys. Usp. **7**, 571 (1965)].

144. E. Brezin and G. Parisi, J. Phys. C **13**, L307 (1980).

145. J. L. Cardy, J. Phys. C **11**, L321 (1978).

146. M. V. Sadovskii, in *Soviet Science Reviews: Physics*, Ed. by I. M. Khalatnikov (Harwood Academic, New York, 1986), Vol. 7, p. 1.

147. I. M. Suslov, Zh. Éksp. Teor. Fiz. **108**, 1686 (1995) [JETP **81**, 925 (1995)].

Translated by A. Betev