Journal of Experimental and Theoretical Physics, Vol. 101, No. 6, 2005, pp. 1009–1020.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 6, 2005, pp. 1156–1168.

ATOMS, MOLECULES, OPTICS

Ionization of Excited States of the Hydrogen Atom by a Laser Pulse

B. A. Zon and A. S. Kornev

Voronezh State University, Universitetskaya pl. 1, Voronezh, 394006 Russia
e-mail: zon@niif.vsu.ru; a-kornev@yandex.ru

Received June 20, 2005

Abstract—The energy spectra of electrons are calculated in the adiabatic approximation when the excited 2*s*, 2*p*, and 3*d* states of the hydrogen atom are ionized by a superstrong ultrashort laser pulse. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Recently, significant progress has been made in the experimental investigation of the ionization of atoms by a superstrong laser field. In particular, this applies to the detailed measurements of the energy distribution of electrons [1–4], as well as to the dependence of the ionization probability on the phase of the light field that arises in ultrashort laser pulses [5].

New experimental results have stimulated theoretical studies in which the ionization of atoms in a strong laser field is considered both by analytical and numerical methods (see [6], as well as [7–11]). The basic problem of the theory is associated with taking into consideration how the Coulomb field of an atomic residual affects the motion of an electron in the continuum. The point is that, for an electron bound by short-range forces, the ionization theory in a strong field was developed as early as the classical work by Keldysh [12]. The development of this method carried out in [13] (see also [14]) allowed one to obtain the spectra of tunneling electrons that are in good agreement with the experimental data of [15]. The authors of [13] put forward a hypothesis that, after a simple modification, analytical formulas that describe the electron spectra for the photodetachment from negative ions (a short-range potential) can also be applied to describe the ionization of neutral atoms and positive ions with long-range Coulomb interaction. The verification of this hypotheses is one of the goals of the present paper.

Analytical methods for the description of electron spectra under the ionization of neutral atoms and positive ions were also developed in [16, 17]. In [17], the author applied the Volkov functions with a Coulomb correction to describe the motion of a free electron in the Coulomb field and the field of an electromagnetic field; however, [18] cast serious doubt on the accuracy of these functions in the case of a strong field.

In the present work, we carried out the calculations on the basis of adiabatic approximation that we proposed and tested earlier in [19]. Recall that the basic idea of the adiabatic approximation used is based on the results of [20]. In that paper, the probability, found by Keldysh, of the tunneling detachment of an electron from a short-range potential was generalized to the case of the tunneling ionization of an atom. This was done by formally replacing the strength of the electric field in the probability of tunneling in a dc field [21] by the strength of an ac electric field of the light wave followed by the averaging of this probability over the field cycle. In the adiabatic approximation [19], the ac field is replaced by a dc field at an earlier stage, in the electron wavefunctions, which can be used for calculating various quantities, e.g., the electron spectra in the present case. After that, the quantities obtained are averaged over the period of the field. It is obvious that the accuracy of such an approach must be no less than the accuracy of calculating the total probability of tunneling [22]. Other variants of the adiabatic approximation were considered in [23].

In this paper, we use the atomic system of units ($\hbar = e = m = 1$).

2. WAVEFUNCTIONS IN CONTINUUM

Consider an electron that moves in the Coulomb potential due to a charge Z and in a laser field of strength $F(t)$ that is linearly polarized along axis z. The wavefunction of the electron in the dipole approximation satisfies the Schrödinger equation

$$i\frac{\partial}{\partial t}\Psi(\mathbf{r}, t) = \left[-\frac{1}{2}\nabla^2 - \frac{Z}{r} + zF(t)\right]\Psi(\mathbf{r}, t). \qquad (1)$$

Let us construct a solution to the Schrödinger equation (1) that is valid in the case of low frequencies ω of the field. We restrict consideration to the case of the

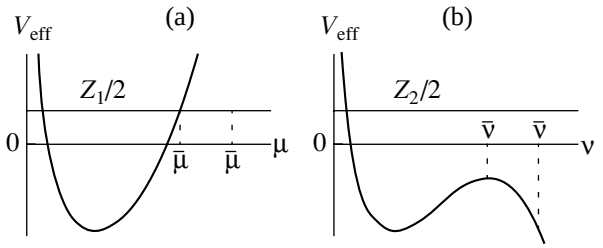


Fig. 1. Effective potentials for the one-dimensional Schrödinger equations (7) (a) and (8) (b).

continuum. We will use the adiabatic approximation, according to which the motion of an electron is adjusted to the variation of the field. Physically, this means that we neglect the inertial properties of the electron when it interacts with the field. According to the classical law of variation of momentum, the applicability condition of the adiabatic approximation in a continuous spectrum can be formulated as follows:

$$\frac{F}{\omega} \gg \sqrt{E}, \tag{2}$$

where F is the characteristic value of the field strength and E is the characteristic value of the electron energy. Condition (2) improves the applicability condition of the adiabatic approximation that was formulated in [19]: $F/\omega \ll 1$. Formula (2) is analogous to the relevant condition that was first proposed in [16].

We write an adiabatic solution to Eq. (1) as follows:

$$\Psi(\mathbf{r}, t) = \Phi_E(F, \mathbf{r})|_{F = F(t)} \exp(-iEt), \tag{3}$$

where $\Phi_E(F, \mathbf{r})$ is a solution to the stationary Schrödinger equation in a uniform dc field F:

$$\left[-\frac{1}{2}\nabla^2 - \frac{Z}{r} + zF\right]\Phi_E(F, \mathbf{r}) = E\Phi_E(F, \mathbf{r}). \tag{4}$$

One can see that function (3) and energy E depend on time parametrically; this is the main approximation of the present work.

It is convenient to solve stationary Schrödinger equation (4) in squared parabolic coordinates μ, ν, φ [24] that are related to the Cartesian coordinates by the formulas

$$x = \mu\nu\cos\varphi, \quad y = \mu\nu\sin\varphi, \quad z = \frac{1}{2}(\mu^2 - \nu^2). \tag{5}$$

Coordinates (5) are related to the conventional parabolic coordinates (ξ, η) by simple formulas $\mu^2 = \xi$ and $\nu^2 = \eta$. Squared coordinates are convenient for the numerical solution of Schrödinger equation (4) because the integration domain becomes more compact [25]. The main properties of the squared parabolic coordinates are summarized in Appendix A.

In these coordinates, the variables in Eq. (4) are separated,

$$\Phi_E(F, \mathbf{r}) = \frac{1}{\sqrt{2\pi\mu\nu}} M(F, \mu) N(F, \nu) e^{im\varphi}, \tag{6}$$

and lead to two one-dimensional Schrödinger equations,

$$\left[\frac{d^2}{d\mu^2} + \frac{1 - 4m^2}{4\mu^2} + 2E\mu^2 - F\mu^4 + Z_1\right]M(F, \mu) = 0, \tag{7}$$

$$\left[\frac{d^2}{d\nu^2} + \frac{1 - 4m^2}{4\nu^2} + 2E\nu^2 + F\nu^4 + Z_2\right]N(F, \nu) = 0, \tag{8}$$

in which the separation constants Z_1 and Z_2 are related by

$$Z_1 + Z_2 = 4Z.$$

Equation (7) describes a finite motion, and its eigenfunctions are characterized by a parabolic quantum number n_1. Equation (8) describes an infinite motion, and its eigenfunctions can be characterized by the total energy E (Fig. 1). Thus, the adiabatic solution to time-dependent Schrödinger equation (1) can be expressed as a parabolic wave

$$\Psi_{En_1m}(\mathbf{r}, t) = \frac{1}{\sqrt{2\pi\mu\nu}} M_{En_1m}(F, \mu) N_{En_1m}(F, \nu)|_{F = F(t)} \times \exp[i(m\varphi - Et)]. \tag{9}$$

A detailed description of the algorithm used for the numerical solution of Eqs. (7) and (8) is presented in Appendix B.

3. AMPLITUDE AND PROBABILITY OF THE PROCESS

Let us choose a laser pulse in the form

$$F(t) = F_0\cos^2\left(\frac{\pi t}{2T}\right)\cos(\omega t - \Theta), \quad |t| \le T, \tag{10}$$

where Θ is the phase of the light field, which is essential for short pulses, and T is the FWHM of the pulse.

The differential (with respect to energy) probability of the bound–free transition between the initial $|i\rangle$ and final $|f\rangle$ states is given by

$$\frac{dP_{fi}}{dE} = \left|\int_{-T}^{T} A_{fi}(t)dt\right|^2, \tag{11}$$

where the wavefunction of the final state is assumed to be normalized to the δ function over the energy scale. The transition probability is

$$A_{fi}(t) = F(t)\langle f|z|i\rangle. \quad (12)$$

Thus, in contrast to the well-known study by Keldysh [1, 2], who chose the Gordon–Volkov function that describes the motion of an electron in a plane electromagnetic field as the wavefunction of the final state, here we choose adiabatic functions that turn, as $\omega \longrightarrow 0$, into an exact solution to the Schrödinger equation that describes the motion of an electron in the Coulomb and dc electric fields.

The wavefunction of the initial state in squared parabolic coordinates is expressed as

$$|i\rangle = |n_{1i}n_{2i}m\rangle = \frac{Z\exp[i(m\varphi - E_0 t)]}{\sqrt{\pi n^3 \mu\nu}} \times f_{n_{1i}|m|}\left(\mu\sqrt{\frac{Z}{n}}\right) f_{n2i|m|}\left(\nu\sqrt{\frac{Z}{n}}\right). \quad (13)$$

Here,

$$n = n_{1i} + n_{2i} + |m| + 1,$$

$$f_{n'm}(x) = \frac{1}{m!}\sqrt{\frac{(n'+m)!}{n'!}} \times x^{(2m+1)/2}\,{}_1F_1(-n', m+1; x^2)\exp\left(-\frac{x^2}{2}\right), \quad (14)$$

${}_1F_1$ is a degenerate hypergeometric function, and $E_0 = -Z^2/2n^2$ is the energy of the bound state.

After the substitution of (9) and (13) into (12), the expression for the amplitude of the process is rewritten as

$$A^{(m)}_{E, n_{1i}n_{2i}}(n_1) = \int_{-T}^{T} \mathcal{N}^{(m)}_{E, n_{1i}n_{2i}}(n_1, t)\exp[i(E - E_0)t]dt, \quad (15)$$

where

$$\mathcal{N}^{(m)}_{E, n_{1i}n_{2i}}(n_1, t) = \frac{ZF}{\sqrt{2n^3}}[\mathcal{J}_4\mathcal{J}_0 - \mathcal{J}_0\mathcal{J}_4]\Big|_{F = F(t)}, \quad (16)$$

$$\mathcal{J}_k = \int_0^\infty \mu^k M_{En_1|m|}(|F|, \mu) f_{n_{1i}|m|}\left(\mu\sqrt{\frac{Z}{n}}\right)d\mu,$$

and $\mathcal{J}_k$ differs from $\mathcal{J}_k$ by the replacements $\mu \longrightarrow \nu$, $M_{En_1|m|} \longrightarrow N_{En_1|m|}$, and $n_{1i} \longrightarrow n_{2i}$. Thus, the ionization amplitude of the bound state is the Fourier image of function (16) with respect to time at the frequency $E - E_0$:

$$A^{(m)}_{E, n_{1i}n_{2i}}(n_1) = \mathcal{N}^{(m)}_{E, n_{1i}n_{2i}}(n_1, \Omega)\big|_{\Omega = E - E_0}. \quad (17)$$

Strictly speaking, expression (16) is valid only for positive half-cycles of the laser field. When integrating in (11) over negative half-cycles, one should make an additional change $n_{1i} \rightleftharpoons n_{2i}$ on the right-hand side of (16) because the coordinates μ and ν exchange their roles for $F < 0$. Recall that the projection of the orbital momentum of an electron onto the direction of linear polarization is preserved; therefore, amplitude (17) depends on m parametrically. Note that, generally speaking, the adiabatic approximation cannot be applied at the moments when the field $F(t)$ changes its sign. However, this restriction does not significantly tells on the amplitude in view of the common multiplier $F(t)$ in (12).

According to (11), the differential (with respect to energy) probability of emission of an electron with a parabolic quantum number n_1 is obtained by squaring (17):

$$\frac{d}{dE}P^{(m)}_{E, n_{1i}n_{2i}}(n_1) = \left|A^{(m)}_{E, n_{1i}n_{2i}}(n_1)\right|^2. \quad (18)$$

The energy distribution of electrons integrated over angles is obtained by summing the probabilities (18) over all values of the parabolic quantum number n_1:

$$\frac{d}{dE}P^{(m)}_{E, n_{1i}n_{2i}} = \sum_{n_1 = 0}^{\infty}\left|A^{(m)}_{E, n_{1i}n_{2i}}(n_1)\right|^2. \quad (19)$$

Thus, the use of parabolic coordinates does not require the solution of a system of coupled differential equations. One should only carry out a summation over the contributions of all parabolic partial waves.

If the initial state is defined by the principal n_i and the orbital l_i quantum numbers, one should take as the initial state $|i\rangle$ an appropriate superposition of states (13) with parabolic numbers n_{1i} and n_{2i} (see, for example, [26]):

$$|i\rangle = |n_i l_i m\rangle = \sum_{n_{1i}, n_{2i}} C^{l_i m}_{\frac{n_i - 1}{2}\frac{m + n_{1i} - n_{2i}}{2}\frac{n_i - 1}{2}\frac{m + n_{2i} - n_{1i}}{2}}|n_{1i}n_{2i}m\rangle. \quad (20)$$

The summation indices in (20) run over nonnegative integers that satisfy condition (14) and do not allow the Clebsch–Gordan coefficient to vanish identically.

4. DISCUSSION OF THE RESULTS

All the numerical results obtained in this paper refer to the energy dependence of the emission probability of

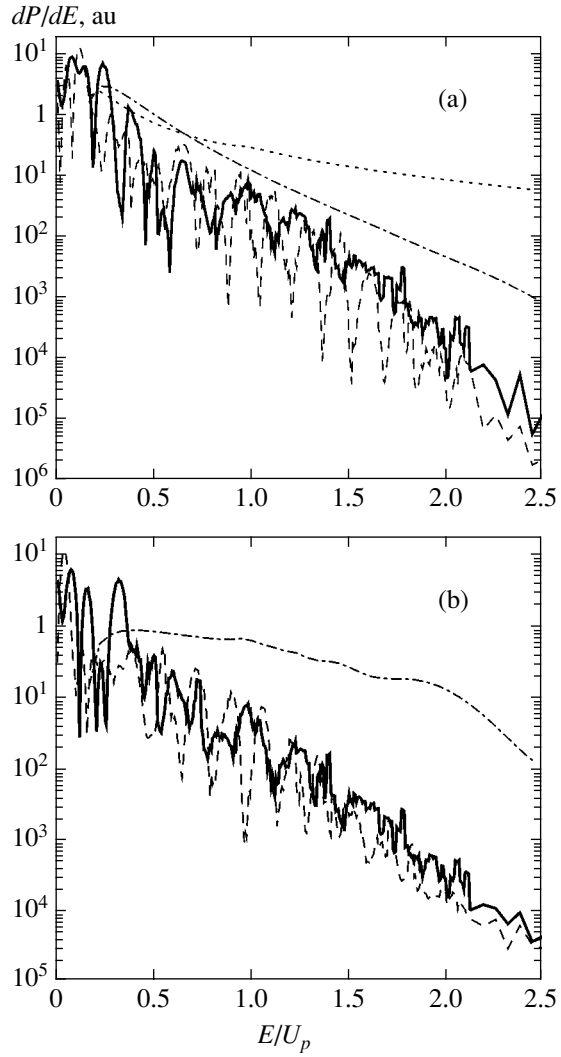


Fig. 2. The spectra of electrons when the hydrogen atom is ionized from the 2*s* (a) and 2*p* (b) states with $m = 0$. Solid lines correspond to a cosine-shaped pulse; dashed lines, to a sine-shaped pulse; dot-and-dash lines represent the results of the kinetic model (25) based on the data of [13]; and the dotted line represents the results of the kinetic model based on the results of [17]. Radiation parameters (10) are as follows: $\omega = 0.056$ au ($\lambda = 800$ nm), the peak intensity is 3.45×10^{14} W/cm^2 ($F_0 = 0.1$ au), and $T = 2.5$ fs.

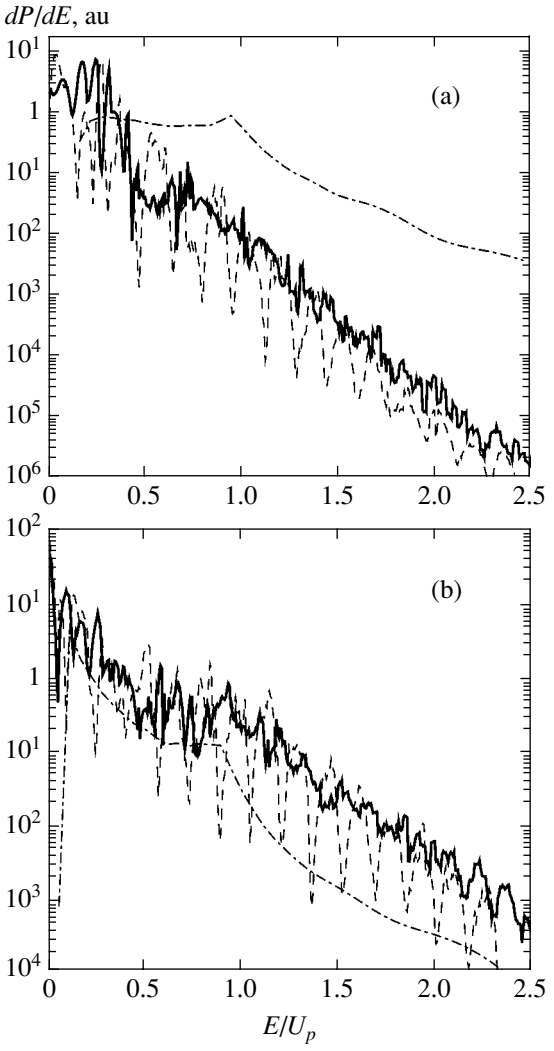


Fig. 3. The spectra of electrons when the hydrogen atom is ionized from the 2*p* state with $|m| = 1$ (a) and from the 3*d* state with $|m| = 2$ (b). The notation and the parameters are the same as those in Fig. 2.

electron integrated over angles. First, note that the adiabatic approximation considered does not allow us to compare our results with those of the studies [27–29], in which the authors carried out analogous calculations for the ground state of the hydrogen atom. The laser parameters used in those papers satisfy adiabaticity condition (2) only for low energies. The energy spectrum of an electron turns out to be rather wide. It appreciably decreases only for energies that fall outside adiabaticity condition (2).

Therefore, we carried out calculations for the ionization of excited (2*s*, 2*p*, and 3*d* with $|m| = 2$) states of the hydrogen atom for $\omega = 0.056$, $F = 0.1$, and $T = 2.5$ fs. The electron spectra for the ionization from excited states are narrower than those for the ionization from the ground state, and the electron yield decreases by several orders of magnitude for the energies that satisfy the adiabaticity condition (2). Moreover, for excited states, the energy of an electron in a field with a maximal strength of $F = 0.1$ is greater than the binding energy of the electron in atom; therefore, formally, the results presented correspond to the case of barrier-suppressed ionization. The question concerning the ionization from the excited states of an atom was repeatedly raised in the literature (see, for example, [4, 30]).

The results of calculations are represented in Figs. 2 and 3; the electron energy is measured in the units of ponderomotive energy $U_p = F_0^2/4\omega^2$.

The results are compared with analogous results obtained in the models considered in [13, 17]. In these

works, analytic formulas were presented for the ionization rates of an atom by linearly polarized monochromatic radiation. In these formulas, the ionization rates naturally depend on the number of absorbed photons. In both models, the quantity W_n was obtained by the numerical integration of the differential (with respect to angles) ionization rate:

$$W_n = 2\pi\int_0^\pi \frac{dW_n}{d\Omega}\sin\theta d\theta.$$

The form of the latter formula depends on the model used.

In [13], the initial state of an electron was taken from the model of a short-range potential, the final state of the electron was described by the Gordon–Volkov wavefunction, and the calculations were carried out in the semiclassical approximation. The result for linearly polarized radiation has the form

$$\frac{dW_n}{d\Omega} = \frac{pA^2}{4\pi}\left(\frac{\kappa}{\omega}\right)^{2\nu} 2\nu \times \Gamma\left(1+\frac{\nu}{2}\right)(2l+1)\frac{(l-|m|)!}{(l+|m|)!} \times \left|P_l^{|m|}\left(\sqrt{1+\frac{p^2\sin^2\theta}{\kappa^2}}\right)\right|^2 \times \left|\sum_{\mu=1,2}(\pm 1)^{l+m}\frac{(c_\mu + is_\mu)^n}{\sqrt{2\pi(-iS''_\mu)^{\nu+1}}}\exp[-ic_\mu(\xi+zs_\mu)]\right|^2. \quad (21)$$

Here, the upper and lower signs correspond to $\mu = 1$ and $\mu = 2$, respectively;

$$s_\mu = -\frac{\xi \pm i\sqrt{8z(n-z)-\xi^2}}{4z},$$
$$c_\mu = \pm\sqrt{1-s_\mu^2},$$
$$S''_\mu = c_\mu(\xi + 4zs_\mu), \quad z = \frac{F^2}{4\omega^3}, \quad (22)$$
$$p = \sqrt{2\left(n\omega - \frac{F^2}{4\omega^2} - E_0\right)},$$

P is the Legendre polynomial; $\xi = \mathbf{F}\cdot\mathbf{p}/\omega^2$; E_0 is the binding energy of the electron in the initial state; and l and m are the orbital and magnetic quantum numbers of the electron, respectively. The constants A, ν, and $\kappa = \sqrt{2E_0}$ are determined by the wavefunction of the initial state

$$\Phi_0(\mathbf{r}) = Ar^{\nu-1}e^{-\kappa r}Y_{lm}(\theta,\varphi), \quad (23)$$

for a short-range potential, $\nu = 0$.

For a neutral atom or a positive ion, formula (23) with $\nu = 1$ provides a correct asymptotic description of the wavefunction with appropriate replacements of the normalization factor A and the binding energy E_0; therefore, an assumption was made in [13] that formula (21) can also be applied to describe the ionization electron spectra of neutral atoms, rather than solely for negative ions. However, the neglect of the effect of the Coulomb field of an atomic residual on the motion of a free electron implies the neglect of rescattering phenomena that have been observed in the recent work [31]; from the formal point of view, this is not admissible in the case of a strong field.

In [17], the Gordon–Volkov wavefunction with a Coulomb correction was used for describing the final state of an electron. Here, we present a result for the barrier-suppressed ionization of the s state of a hydrogen-like atom with the charge number Z and the principal quantum number n_p by linearly polarized radiation:

$$\frac{dW_n}{d\Omega} = \frac{p^2\omega^2 ZD^2}{\pi^2(2F)^{4/3}n_p^2}\mathrm{Ai}^2\left(\frac{Z^2/n_p^2 + p_\perp^2 + p_\parallel^2\gamma^2/3}{(2F)^{2/3}}\right). \quad (24)$$

Here,

$$D = \left(\frac{4Z^3}{Fn_p^4}\right)^{n_p},$$

$\gamma = \omega Z/Fn_p$ is the Keldysh parameter; $p_\parallel$ and $p_\perp$ are the longitudinal and transverse (with respect to the polarization vector) components of momentum, respectively; and Ai is the Airy function. Paper [17] does not contain analytic expressions for the ionization rate of the states with $l > 0$.

The rates (21) and (24) of ionization from the excited states of the hydrogen atom prove to be so large that their product multiplied by a time interval on the order of the laser-pulse duration yields a probability greater than 1. Therefore, in the present case, these quantities should be treated precisely as transition rates rather than transition probabilities per unit time.

To obtain reasonable results with the rates (21) and (24), we wrote kinetic equations that were solved for a pulse with the envelope from formula (10) (unfortu-

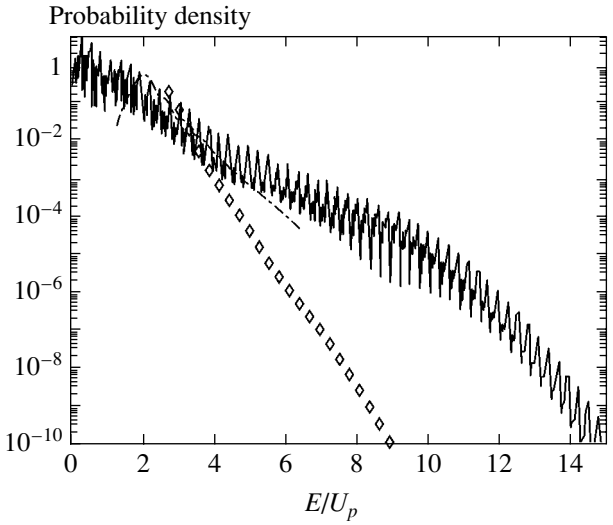


Fig. 4. Electron yield under the tunneling ionization of the hydrogen atom from the 1*s* state. Comparison of the results of [29] (solid line) with those of the model of [13] (rhombs) and the kinetic model (25) based on [13] (dot-and-dash line). Radiation parameters are as follows: $\omega = 0.074$ au ($\lambda = 680$ nm), and the peak intensity is 2×10^{14} W/cm^2 ($F_0 = 0.076$ au).

nately, the dependence on the initial phase of the pulse in this case is lost):

$$\frac{dC_0}{dt} = -\sum_{n'=1}^{N_{\max}} W_{n'}C_0, \quad \frac{dC_n}{dt} = W_nC_0, \qquad (25)$$
$$C_0(-T) = 1, \quad C_n(-T) = 0, \quad n = 1, \ldots, N_{\max}.$$

Here, C_n is the probability of ionization with the absorption of n photons, C_0 is the initial concentration of neutral atoms, and $N_{\max}$ is the maximal number of absorbed photons taken into account in the calculations.

System (25) can formally be solved in quadratures, which, however, cannot be applied to numerical calculations in view of the insufficiently smooth behavior of the integrands. Therefore, it is more convenient to solve numerically the kinetic equations themselves. Unfortunately, this problem is stiff [32]. For such systems, special numerical methods were developed in [33].

The solution of system (25) results in a set of ionization probabilities that differ by the number of absorbed photons. Then, these probabilities are interpolated over a continuous spectrum of energies. The values of the probabilities contain at most seven correct digits; this restricts $N_{\max}$ and, hence, the upper boundary of the energy spectra of electrons.

As additional information, Fig. 4 presents a comparison of the electron spectra obtained in [29], when the hydrogen atom is ionized from the ground state, with the results obtained in [13] by multiplying the ionization rate by the length of the laser pulse. One can see that the spectrum obtained in [13] has a substantially steeper slope compared with the results of [29]; apparently, this is associated with the neglect of rescattering processes. The results of [29] cannot be compared with the model of [17] because, for the radiation parameters chosen in [29], the Keldysh parameter is not sufficiently small.

The diagrams in Fig. 4 show that the electron spectra do not contain plateaulike regions in which the electron distribution relatively weakly depends on energy. This conclusion is in agreement with the results of [34], in which the authors established theoretically and experimentally that such plateaulike regions disappear as the intensity of laser radiation intensity increases.

5. CONCLUSIONS

The main results obtained in the present paper are as follows.

1. Within an adiabatic model, we have calculated the energy spectra of electrons generated as a result of ionization of hydrogen atoms from excited states by an intense laser pulse.

2. A kinetic model based on the analytic formulas of [13] well describes the spectra of electrons in the low-energy domain; however, in the high-energy region, the agreement between the results of this model and our results is worse. This fact by no means implies that the model of [13] is inapplicable in the general case, because our calculations were performed for a very short laser pulse.

3. The same conclusion applies to the model proposed in [17]. The results of this model can be compared with the results of our calculations only in the case of ionization of the hydrogen atom from the 2*s* state.

ACKNOWLEDGMENTS

We are grateful to M.Yu. Kuchiev for useful remarks. This work was performed within the departmental research program “Development of the Scientific Potential of the Higher School” (code no. 15286) under the support of the Russian Foundation for Basic Research (project no. 05-02-16253), the Council on Grants of the President of the Russian Federation (project no. MK-3551.2004.2), the US Civilian Research and Development Foundation (project no. VZ-010-0), and the Dynasty Foundation.

APPENDICES

Appendix A

Squared Parabolic Coordinates

Here, we present the form of the scale factors (see, for example, [24]) for squared parabolic coordinates (5):

$$h_\mu = h_\nu = \sqrt{\mu^2 + \nu^2}, \quad h_\varphi = \mu\nu. \qquad (A.1)$$

Using these factors, we can construct all vector differential operators as applied to the given system of coordinates. For example, the volume element has the form

$$d^3r = \mu\nu(\mu^2 + \nu^2)d\mu d\nu d\varphi. \quad \text{(A.2)}$$

The components of the operator ∇ are expressed as

$$\nabla_\mu = \frac{1}{\sqrt{\mu^2 + \nu^2}}\frac{\partial}{\partial\mu}, \quad \nabla_\nu = \frac{1}{\sqrt{\mu^2 + \nu^2}}\frac{\partial}{\partial\nu}, \quad \nabla_\varphi = \frac{1}{\mu\nu}\frac{\partial}{\partial\varphi}. \quad \text{(A.3)}$$

Finally, the Laplacian in the variables (μ, ν, φ) has the following structure:

$$\nabla^2 = \frac{1}{\mu\sqrt{\mu^2 + \nu^2}}\frac{\partial}{\partial\mu}\left(\frac{\mu}{\sqrt{\mu^2 + \nu^2}}\frac{\partial}{\partial\mu}\right) + \frac{1}{\nu\sqrt{\mu^2 + \nu^2}}\frac{\partial}{\partial\nu}\left(\frac{\nu}{\sqrt{\mu^2 + \nu^2}}\frac{\partial}{\partial\nu}\right) + \frac{1}{\mu^2\nu^2}\frac{\partial^2}{\partial\varphi^2}. \quad \text{(A.4)}$$

Appendix B

General Description of the Algorithm

Formally, Eq. (7) corresponds to a finite one-dimensional motion with "energy" $Z_1/2$ (Fig. 1a). The boundary conditions for $M(E, \mu)$ are given by

$$M(F, \mu)|_{\mu \to 0} \sim \mu^{(2|m| + 1)/2}, \quad \text{(B.1)}$$

$$M(F, \mu)|_{\mu \to \infty} \sim \frac{1}{\mu}\exp\left(-\frac{F^{1/2}}{3}\mu^3 + \frac{F}{F^{1/2}}\mu\right). \quad \text{(B.2)}$$

Let $\tilde{\mu}$ be the first classical turning point. A convenient method for solving Eq. (7) is given by the step-by-step expansion of $M(F, \mu)$ in power series [25]. To this end, the function $M(F, \mu)$ is first expanded in a power series

$$M(F, \mu) = \mu^{(2|m| + 1)/2}\sum_{p=0}^{\infty} C_{0p}\mu^{2p} \quad \text{(B.3)}$$

in the neighborhood of zero up to the point $\mu = \mu_1 = 1$ and then is expanded in power series in the neighborhoods of other points $\mu_1 < \mu_i \le \tilde{\mu}$ that lie at a distance of 0.5 from each other:

$$M(F, \mu) = \sum_{p=0}^{\infty} C_{ip}(\mu - \mu_i)^p; \quad \text{(B.4)}$$

i.e., in each interval $[\mu_k, \mu_{k+1}]$, a separate expansion of the function $M(F, \mu)$ is used. The method of calculating the coefficients C_{ip} is described below. The summation of series (B.3) and (B.4) is terminated when the absolute values of three consecutive terms of the series are less than the maximal of these values by a factor of 10^{12}. This method, which was proposed in [25], is convenient since it does not require the partition of the integration interval into a large number of subintervals and thereby significantly speeds up the calculations. The oscillations of the function $M(F, \mu)$ do not produce any appreciable effect on the stability of the algorithm.

Nevertheless, in the classically inaccessible domain $(\mu > \tilde{\mu})$, this method leads to incorrect results due to the effect of the exponentially growing solution of (7) that violates the boundary condition (B.2). To overcome this drawback, we performed a standard numerical integration of Eq. (7) in the classically inaccessible region starting from the point $\tilde{\mu}$ satisfying the condition

$$\frac{1}{\tilde{\mu}}\exp\left(-\frac{F^{1/2}}{3}\tilde{\mu}^3 + \frac{E}{F^{1/2}}\tilde{\mu}\right) < 10^{-6}$$

to the point $\tilde{\mu}$; i.e., we integrated in backwards. The boundary condition was chosen according to (B.2). The separation constant Z_1 was determined from the condition of sewing together the logarithmic derivatives of the function $M(F, \mu)$ on the left and right of the turning point $\tilde{\mu}$:

$$\frac{d}{d\mu}\ln|M(F, \mu)|\Big|_{\mu = \tilde{\mu} - 0} = \frac{d}{d\mu}\ln|M(F, \mu)|\Big|_{\mu = \tilde{\mu} + 0}. \quad \text{(B.5)}$$

The parameter Z_1 takes discrete values numbered by the parabolic quantum number $n_1 = 0, 1, \ldots$, which is equal to the number of zeros of the function $M(F, \mu)$. For a given n_1, it is convenient to seek the quantity Z_{1, n_1} by solving (B.5) for Z_1 in a small neighborhood of its semiclassical value $Z_{1, n_1}^{(\mathrm{WKB})}$ (see below):

$$Z_{1, n_1}^{(\mathrm{WKB})} - \Delta Z \le Z_{1n_1} \le Z_{1, n_1}^{(\mathrm{WKB})} + \Delta Z,$$
$$\Delta Z \approx 0.1[Z_{1, n_1 + 1}^{(\mathrm{WKB})} - Z_{1, n_1}^{(\mathrm{WKB})}].$$

As a rule, the difference between the exact and semiclassical values of Z_{1n_1} is no greater than 1%. This difference decreases as energy E increases. For $E > 1$, the contribution of the classically inaccessible region to the observable characteristics of the ionization process does not exceed 0.5%. In this case, we can set the wavefunction $M(F, \mu)$ equal to zero in the domain of $\mu > \tilde{\mu}$ and restrict the analysis to the semiclassical formulas for the expansion constant Z_1.

Equation (8) formally differs from (7) by the replacements

$$F \longrightarrow -F, \quad Z_1 \longrightarrow Z_2 = 4Z - Z_1. \tag{B.6}$$

For $E > 0$, Eq. (8) corresponds to an infinite one-dimensional motion with energy $Z_2/2$ (see Fig. 1b). Thus, while moving in the direction of ν, an electron passes over the barrier. Denote by $\tilde{\nu}$ the value of the coordinate at which the barrier height attains its maximum:

$$\tilde{\nu}^2 = \frac{1}{3}\left(V - a - \frac{a^2}{V}\right),$$

where

$$V = \frac{1}{2^{1/3}}[\sqrt{b(4a^3 + b)} - 2a^3 - b]^{1/3},$$

$$a = \frac{E}{F}, \quad b = \frac{27}{8F}(4m^2 - 1).$$

The integration of Eq. (8) over the interval $0 \le \nu \le \tilde{\nu}$ reproduces the corresponding procedure with Eq. (7) up to the replacements (B.6). However, the boundary condition for $\nu \longrightarrow \infty$ differs from (B.2); namely,

$$N(F, \nu)|_{\nu \to \infty} \approx \frac{A}{\nu}\sin\left(\frac{F^{1/2}}{3}\nu^3 + \frac{E}{F^{1/2}}\nu + \phi_0\right), \tag{B.7}$$

where A is a normalization constant and ϕ_0 is the phase calculated during the integration.

During the numerical integration of Eq. (8), the asymptotic representation (B.7) proves to be virtually inaccessible, because it becomes valid only for $\nu \sim 1000$. The phase ϕ_0 was calculated with the use of an improved semiclassical approximation for the function $N(F, \nu)$ (see below):

$$N_{\mathrm{WKB}}(F, \nu) = \frac{D}{\sqrt{X(\nu)}}\sin\left[\int^{\nu} X(\nu')d\nu' + \beta\right], \tag{B.8}$$

$$X^2(\nu) = k^2(\nu) - \frac{1}{4k^2(\nu)}\frac{d^2}{d\nu^2}k^2(\nu) + \frac{5}{16k^4(\nu)}\left[\frac{d}{d\nu}k^2(\nu)\right]^2, \tag{B.9}$$

$$k^2(\nu) = \frac{1 - 4m^2}{4\nu^2} + 2E\nu^2 + F\nu^4 + Z_2. \tag{B.10}$$

Here, β is a constant phase determined by the lower limit of the phase integral and $k^2(\nu)$ makes the sense of the square of the classical momentum. It can be shown that, for $\nu \longrightarrow \infty$, expressions (B.8)–(B.10) imply formula (B.7). In the domain of $\nu > \tilde{\nu}$, the numerical integration of Eq. (8) is continued with a variable step,

$$\Delta\nu_i \equiv \nu_i - \nu_{i-1} = \frac{\pi}{X(\nu_{i-1})},$$

which allows one to avoid a large number of oscillations at each step. Introducing the notation

$$a_1 = N(F, \nu_{i-1})\sqrt{X(\nu_{i-1})},$$

$$a_2 = N(F, \nu_i)\sqrt{X(\nu_i)},$$

$$\alpha = \int_{\nu_{i-1}}^{\nu_i} X(\nu)d\nu,$$

we obtain the following expression for the constant D in the function (B.8):

$$D^2 = (a_1^2 + a_2^2 - 2a_1a_2\cos\alpha). \tag{B.11}$$

The numerical integration of Eq. (8) is performed until the difference between the values of D at two consecutive steps becomes less than 10^{-6}–10^{-7}. The value of $\nu = \bar{\nu}$ at which the integration is terminated is comparable with the value of $\bar{\mu}$ in magnitude. Thus, the numerical solution to Eq. (8) is correctly brought to the required asymptotic form (B.7) by using the semiclassical approximation (B.8). The constant D in (B.8) is related to the constant A in (B.7) by the formula $D = AF^{1/4}$.

It is convenient to normalize the states of the continuous spectrum (6) as follows:

$$\int \Phi'_{E'n_1'm'}(F, \mathbf{r})\Phi_{En_1m}(F, \mathbf{r})d^3r = \delta_{m'm}\delta_{n_1'n_1}\delta(E' - E). \tag{B.12}$$

The possibility of normalizing to the δ function over the energy scale is guaranteed by the infinite motion in the direction of ν. In this case, the main contribution to the integral with respect to ν is made by the domain of $\nu \gg 1$, and condition (B.12) is simplified:

$$\int_0^{\infty} M^*_{En_1'm}(F, \mu)M_{En_1m}(F, \mu)d\mu = \delta_{n_1'n_1}, \tag{B.13}$$

$$\int_0^{\infty} \nu^2 N^*_{E'n_1m}(F, \nu)N_{En_1m}(F, \nu)d\nu = \delta(E' - E). \tag{B.14}$$

The normalization constant of the function $N(F, \nu)$ is chosen according to its asymptotic behavior with the use of (B.14). For large values of ν, the function $N(F, \nu)$ can be expressed in terms of a linear combination of outgoing N_{out} and incoming N_{in} parabolic waves. The normalization to the δ function over the scale of energies in atomic units is equivalent to the normalization to the value of the flux in the outgoing wave $I_{\text{out}} = 1/2\pi$ [35]. Under condition (B.13), this assertion is equivalent to the requirement that $D = \sqrt{2/\pi}$.

When $\nu > \bar{\nu}$, one can apply the semiclassical form of the wavefunction (B8). Its convenient form for calculations is given below.

Appendix C

Calculation of Coefficients in the Expansions of the Function $M(F, \mu)$

The substitution of (B.3) and (B.4) into (7) leads to the following recurrence relations for the coefficients C_{in}:

$$C_{0n} = -\frac{Z_1 C_{0,n-1} + 2E C_{0,n-2} - F C_{0,n-3}}{4n(n+m)}, \qquad \text{(C.1)}$$

$$\begin{aligned} C_{\text{in}} = -\frac{1}{n(n-1)\mu_i^2}\Bigg\{ & 2(n-1)(n-2)\mu_i C_{i,n-1} \\ & + \left[(n-2)(n-3) + \frac{1}{4} - m^2 + Z_1\mu_i^2 + 2E\mu_i^4 - F\mu_i^6\right] \\ & \times C_{i,n-2} + 2(Z_1\mu_i + 4E\mu_i^3 - 3F\mu_i^5)C_{i,n-3} \\ & + (Z_1 + 12E\mu_i^2 - 15F\mu_i^4)C_{i,n-4} \\ & + 4(2E\mu_i - 5F\mu_i^3)C_{i,n-5} + (2E - 15F\mu_i^2)C_{i,n-6} \\ & - 6F\mu_i C_{i,n-7} - F C_{i,n-8}\Bigg\}. \end{aligned} \qquad \text{(C.2)}$$

According to the boundary condition (B.1), we must set

$$C_{0,-1} = C_{0,-2} = C_{0,-3} = 0$$

and choose the constant (B.1) with a nonzero initial value. The initial values of C_{in} ($i > 0$) are calculated from the condition of sewing together the series (B.3) and (B.4) at the boundaries of their domains of definition. The sewing consists in equating the functions and their first and second derivatives to the left and right of the boundary point μ_i. The second derivatives are equated so that the one-dimensional Schrödinger equation (7) hold at the boundary point μ_i itself. Here, we present the final results. For $i = 1$, we have

$$C_{10} = \mu_1^{(2|m|+1)/2}\sum_{n=0}^{\infty} C_{0n}\mu_1^{2n},$$

$$C_{11} = \mu_1^{(2|m|+1)/2}\sum_{n=0}^{\infty}\left[2n + |m| + \frac{1}{2}\right]C_{0n}\mu_1^{2n-1},$$

$$C_{12} = \frac{1}{2}\mu_1^{(2|m|+1)/2}\sum_{n=0}^{\infty}\left[(2n+|m|)^2 - \frac{1}{4}\right]C_{0n}\mu_1^{2n-2},$$

$$C_{1,-1} = \ldots = C_{1,-5} = 0.$$

For $i > 1$, we have

$$C_{i,0} = \sum_{n=0}^{\infty} C_{i-1,n}(\Delta\mu_i)^n,$$

$$C_{i,1} = \sum_{n=1}^{\infty} nC_{i-1,n}(\Delta\mu_i)^n,$$

$$C_{i,2} = \frac{1}{2}\sum_{n=2}^{\infty} n(n-1)C_{i-1,n}(\Delta\mu_i)^n,$$

$$C_{i,-1} = \ldots = C_{i,-5} = 0.$$

Here, $\Delta\mu_i = \mu_i - \mu_{i-1}$.

Appendix D

Semiclassical Formula for the Separation Constant

To calculate the separation constant Z_1 for $E > 0$, a semiclassical formula was derived in [36]. Here, we present it as applied to squared parabolic coordinates.

Introduce the notation $k = \sqrt{2E}$. For

$$k^3 < 3\pi\left[n_1 + \frac{1}{2}(|m|+1)\right]F$$

we have

$$Z_{1,n_1}^{(\text{WKB})} = \frac{t}{(t-1)^2}\frac{k^4}{F},$$

where the parameter $t > 1$ is determined from the equation

$$\frac{t}{(t-1)^{3/2}}{}_2F_1\left(-\frac{1}{2}, \frac{1}{2}; 2; -t\right) = 4\left[n_1 + \frac{1}{2}(|m|+1)\right]\frac{F}{k^3},$$

$$n_1 = 0, 1, \ldots.$$

For

$$k^3 > 3\pi\left[n_1 + \frac{1}{2}(|m| + 1)\right]F$$

we have

$$Z_{1,n_1}^{(\mathrm{WKB})} = -\frac{t}{(t+1)^2}\frac{k^4}{F},$$

$$\frac{(t-1)^2}{(t+1)^{3/2}}{}_2F_1\left(\frac{1}{2}, \frac{3}{2}; 3; 1-t\right)$$

$$= 16\left[n_1 + \frac{1}{2}(|m| + 1)\right]\frac{F}{k^3}.$$

The spectrum Z_1 is bounded from below: $Z_1 \geq -E^2/F$.

Appendix E

Improved Semiclassical Approximation

A solution to the equation

$$\frac{d^2N}{d\nu^2} + W(\nu)N(\nu) = 0, \tag{E.1}$$

where $W(\nu) > 0$, is sought for in the semiclassical approximation as

$$N(\nu) = \frac{D}{\sqrt{X(\nu)}}\sin\left[\int^{\nu} X(\nu')d\nu' + \beta\right]. \tag{E.2}$$

The substitution of (E.2) into (E.1) leads to the following nonlinear equation for the unknown function $X(\nu)$:

$$X^2(\nu) + \frac{1}{2X}\frac{d^2X}{d\nu^2} + \frac{3}{4X^2}\left(\frac{dX}{d\nu}\right)^2 + W = 0. \tag{E.3}$$

If the function $X(\nu)$ varies slowly, all the terms of (E.3) that contain derivatives make a small contribution. Therefore, we use an iterative method for solving Eq. (E.3) by taking $X = \sqrt{W}$ as the zeroth step of iteration. As is known, the traditional semiclassical approximation corresponds to this zeroth iteration step. The first iteration step, i.e., the substitution of $X = \sqrt{W}$ into the derivatives in (E.3), yields

$$X^2(\nu) = W - \frac{1}{4W}\frac{d^2W}{d\nu^2} + \frac{5}{16W^2}\left(\frac{dW}{d\nu}\right)^2. \tag{E.4}$$

As a rule, Eq. (E.3) allows one to reproduce up to eight significant digits obtained by numerically solving Eq. (E.1). The second iteration step yields up to 14 significant digits; however, in this case, the formula for $X(\nu)$ proves to be unstable with respect to the variation of the parameter $W(\nu)$. Setting $W(\nu) = k^2(\nu)$ in (E.4), we arrive at (B.8).

Appendix F

Calculation of the Outgoing Flux

In squared parabolic coordinates, the flux I through the surface S defined by the equation ν = const is expressed as

$$I = \iint_S d\mathbf{S}_\nu \cdot \mathbf{j}, \tag{F.1}$$

$$d\mathbf{S}_\nu = \mu\nu\sqrt{\mu^2 + \nu^2}\,d\mu d\varphi\mathbf{e}_\nu,$$

where $\mathbf{e}_\nu$ is a unit vector in the direction of ν. It is well-known that the current density $\mathbf{j}$ in the state with the wavefunction Ψ is given by

$$\mathbf{j} = \frac{1}{2i}(\Psi^*\nabla\Psi - \Psi\nabla\Psi^*). \tag{F.2}$$

The components ∇ in the squared parabolic coordinates are given by formulas (A.3).

In the state (6) with the asymptotics (B.7), the outgoing component has the form

$$N_{\mathrm{out}}(F, \nu) = \frac{A}{2i}\frac{1}{\nu} \times \exp\left(\frac{i}{3}\sqrt{F}\nu^3 + \frac{iE}{\sqrt{F}}\nu + i\phi_0\right). \tag{F.3}$$

Substituting (6) into (F.2) and (F.1), we obtain

$$I_{\mathrm{out}} = \frac{A^2}{4}\sqrt{F}\int_0^{\infty} M^2(\mu)d\mu. \tag{F.4}$$

Appendix G

Calculation of the Phase Integral

For definiteness, we choose $\bar{\nu}$ as the lower limit in the phase integral of the function (B.8). Then,

$$\phi_0 = \arcsin\left[N(F, \bar{\nu})\sqrt{\frac{\pi X(\bar{\nu})}{2}}\right],$$

where $N(F, \nu)$ is a normalized solution to Eq. (8).

Let us single out analytically the growing terms from the phase integral. To this end, we represent the integrand as

$$X(\nu) = \sqrt{2E\nu^2 + F\nu^4 + \Xi(\nu)},$$

where

$$\Xi(\nu) = \frac{1-4m^2}{4\nu^2} + Z_2 - \frac{q''(\nu)}{4q(\nu)} + \frac{5(q'(\nu))^2}{16q^2(\nu)},$$

$$q(\nu) \equiv k^2(\nu),$$

and transform it by integrating by parts and singling out $\epsilon(\nu) = 2E\nu^2 + F\nu^4$:

$$\int_{\bar\nu}^{\nu} X(\nu')d\nu' = S(\nu) - S(\bar\nu)$$

$$-\int_{\bar\nu}^{\nu}\left(\frac{2}{3}\frac{E}{F} + \frac{\nu'^2}{3}\right)\left\{-\frac{2}{\nu'}z(\nu') + \frac{1}{16}\left[-\frac{4(1-4m^2)}{\nu'^2}\right.\right.$$

$$\left.\left. + \frac{-5q'(\nu')^3 + 7qq'(\nu')q''(\nu') - 2q(\nu')^2q'''(\nu')}{q(\nu')^3}\right]\right\}$$

$$\times\left[1 + \frac{z(\nu')}{\epsilon(z')}\right]^{-1/2}\frac{d\nu'}{\nu'}.$$

Here,

$$S(\nu) = \left(\frac{2}{3}\frac{E}{F} + \frac{\nu^2}{3}\right)\frac{\epsilon(\nu)}{\nu}\sqrt{1 + \frac{\Xi(\nu)}{\epsilon(\nu)}},$$

$$z(\nu) = Z_2 + \frac{1-4m^2}{4\nu^2} + \frac{5q'(\nu)^2 - 4q(\nu)q''(\nu)}{16q(\nu)^2},$$

$$q'(\nu) = -\frac{1-4m^2}{2\nu^3} + 4E\nu + 4F\nu^3,$$

$$q''(\nu) = \frac{3(1-4m^2)}{2\nu^4} + 4E + 12F\nu^2,$$

$$q'''(\nu) = \frac{6(1-4m^2)}{\nu^5} + 24F\nu.$$

REFERENCES

1. B. Walker, B. Sheehy, K. C. Kulander, *et al.*, Phys. Rev. Lett. **77**, 5031 (1996).
2. B. Sheehy, R. Lafon, M. Widmer, *et al.*, Phys. Rev. A **58**, 3942 (1998).
3. B. Buerke and D. D. Meyerhofer, Phys. Rev. A **69**, 051402(R) (2004).
4. V. L. B. de Jesus, B. Feuerstein, K. Zrost, *et al.*, J. Phys. B: At. Mol. Opt. Phys. **37**, L161 (2004).
5. G. G. Paulus, F. Grasbon, H. Walter, *et al.*, Nature **414**, 182 (2001).
6. R. Dörner, T. Weber, M. Weckenbrock, *et al.*, Adv. At. Mol. Opt. Phys. **48**, 1 (2002).
7. A. S. Kornev, E. B. Tulenko, and B. A. Zon, Phys. Rev. A **68**, 043414 (2003).
8. A. S. Kornev, E. B. Tulenko, and B. A. Zon, Phys. Rev. A **69**, 065401 (2004).
9. C. C. Chirila and R. M. Potvliege, Phys. Rev. A **71**, 021402(R) (2005).
10. I. V. Litvinyuk, F. Légaré, P. W. Dooley, *et al.*, Phys. Rev. Lett. **94**, 033003 (2005).
11. V. E. Chernov, I. Yu. Kiyan, H. Helm, and B. A. Zon, Phys. Rev. A **71**, 033410 (2005).
12. L. V. Keldysh, Zh. Éksp. Teor. Fiz. **47**, 1945 (1964) [Sov. Phys. JETP **20**, 1307 (1964)].
13. G. F. Gribakin and M. Yu. Kuchiev, Phys. Rev. A **55**, 3760 (1997).
14. M. V. Frolov, N. L. Manakov, E. A. Pronin, and A. F. Starace, J. Phys. B: At. Mol. Opt. Phys. **36**, L419 (2003).
15. I. Yu. Kiyan and H. Helm, Phys. Rev. Lett. **90**, 183001 (2003).
16. N. B. Delone and V. P. Kraĭnov, J. Opt. Soc. Am. B **8**, 1207 (1991).
17. V. P. Kraĭnov, J. Opt. Soc. Am. B **14**, 425 (1997).
18. A. S. Kornev and B. A. Zon, J. Phys. B: At. Mol. Opt. Phys. **35**, 2451 (2002).
19. B. A. Zon and A. S. Kornev, Zh. Éksp. Teor. Fiz. **123**, 991 (2003) [JETP **96**, 870 (2003)].
20. A. M. Perelomov, V. S. Popov, and M. V. Terent'ev, Zh. Éksp. Teor. Fiz. **50**, 1393 (1966) [Sov. Phys. JETP **23**, 924 (1966)].
21. B. M. Smirnov and M. I. Chibisov, Zh. Éksp. Teor. Fiz. **49**, 841 (1965) [Sov. Phys. JETP **22**, 585 (1966)].
22. M. V. Ammosov, N. B. Delone, and V. P. Kraĭnov, Zh. Éksp. Teor. Fiz. **91**, 2008 (1986) [Sov. Phys. JETP **64**, 1191 (1986)].
23. R. M. Potvliege and R. Shakeshaft, Phys. Rev. A **40**, 3061 (1989); M. Dörr, R. M. Potvliege, and R. Shakeshaft, Phys. Rev. A **41**, 558 (1990); R. M. Potvliege and R. Shakeshaft, Phys. Rev. A **41**, 1609 (1990); R. Shakeshaft, R. M. Potvliege, M. Dörr, and W. E. Cooke, Phys. Rev. A **42**, 1656 (1990).
24. P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953; Inostrannaya Literatura, Moscow, 1958), Chap. 5.
25. R. J. Damburg and V. V. Kolosov, J. Phys. B: At. Mol. Phys. **9**, 3149 (1976).
26. D. Park, Z. Phys. **159**, 155 (1960).

27. S. Dionissopoulou, A. Lyras, Th. Mercouris, and C. A. Nicolaides, J. Phys. B: At. Mol. Opt. Phys. **28**, L109 (1995).

28. S. Dionissopoulou, Th. Mercouris, A. Lyras, and C. A. Nicolaides, Phys. Rev. A **55**, 4397 (1997).

29. E. Cormier and P. Lambropoulos, J. Phys. B: At. Mol. Opt. Phys. **30**, 77 (1997).

30. A. Rudenko, K. Zrost, C. D. Schröter, *et al.*, J. Phys. B: At. Mol. Opt. Phys. **37**, L407 (2004).

31. W. A. Bryan, E. M. L. English, J. Wood, *et al.*, in *Proceedings of 24th International Conference on Photonic, Electronic and Atomic Collisions, XXIV ICPEAC* (Rosario, Argentina, 2005), p. Fr301.

32. *Modern Numerical Methods for Ordinary Differential Equations*, Ed. by G. Hall and J. Watt (Oxford Univ. Press, Oxford, 1976; Mir, Moscow, 1979).

33. L. F. Shampine and H. A. Watts, *DEPAC—Design of a User Oriented Package of ODE Solvers* (Sandia National Laboratory, New Mexico, 1980), Rep. SAND79-2374.

34. G. G. Paulus, W. Nicklich, F. Zacher, *et al.*, J. Phys. B: At. Mol. Opt. Phys. **29**, L249 (1996).

35. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory*, 5th ed. (Fizmatlit, Moscow, 2001; Pergamon, Oxford, 1980), Sects. 37 and 77.

36. V. D. Kondratovich and V. N. Ostrovskiĭ, Zh. Éksp. Teor. Fiz. **79**, 395 (1980) [Sov. Phys. JETP **52**, 198 (1980)].

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 101, No. 6, 2005, pp. 1021–1035.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 6, 2005, pp. 1169–1183.

NUCLEI, PARTICLES, FIELDS, GRAVITATION, AND ASTROPHYSICS

Theory of the Muonic Hydrogen–Muonic Deuterium Isotope Shift

A. P. Martynenko

Samara State University, Samara, 443011 Russia

e-mail: mart@ssu.samara.ru

Received December 16, 2004

Abstract—Corrections of the α^3, α^4, and α^5 orders are calculated for the Lamb shift of the 1*S* and 2*S* energy levels of muonic hydrogen μp and muonic deuterium μd. The nuclear structure effects are taken into account in terms of the charge radii of the proton r_p and deuteron r_d for one-photon interaction, as well as in terms of the electromagnetic form factors of the proton and deuteron for the case of one-loop amplitudes. The μd–μp isotope shift for the 1*S*–2*S* splitting is found to be equal to 101003.3495 meV, which can be treated as a reliable estimate when conducting the corresponding experiment with an accuracy of 10^{-6}. The fine-structure intervals $E(1S) - 8E(2S)$ in muonic hydrogen and muonic deuteron are calculated. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The structure of the energy levels of hydrogen-like atoms is an important test for the Standard Model and can be used to obtain more accurate values of a number of fundamental constants (the fine structure constant, the electron and muon masses, the proton charge radius, etc.) [1–3]. In recent years, considerable interest in this area is associated with the muonic hydrogen atom [4–6]. This interest is primarily stimulated by the intensification of experimental investigations of the 2*P*–2*S* Lamb shift and hyperfine structure of the muonic hydrogen atom. Measurement of the 2*P*–2*S* Lamb shift with an accuracy of 3×10^{-5} will make it possible to obtain the proton charge radius with an accuracy of 10^{-3}, which is by an order of magnitude better than the currently available accuracy known from various sources including electron–proton scattering and the 2*P*–2*S* Lamb shift in the hydrogen atom. Measurement of the hyperfine splitting of the ground state of the muonic hydrogen with a similar accuracy would enable one to determine a new value of another fundamental parameter of the theory, the Zemach radius [7], with an accuracy of 10^{-3} [8, 9]. Then, it would be used to calculate a new theoretical value for hyperfine splitting in the hydrogen atom and to obtain bounds on the proton polarizability correction [10–12].

Finally, there is an additional experimental problem of investigating the 1*S*–2*S* large fine structure in the muonic hydrogen atom and the muonic hydrogen–muonic deuterium isotope shift for this splitting [13, 14], which makes it possible to acquire new data on the charge radii of the proton and deuteron. It is worth noting that both indicated quantities are among the most accurately measured quantities for the hydrogen atom. In particular, the current value of the hydrogen–deuterium isotope shift measured for the 1*S*–2*S* splitting, the measurement accuracy for which increases by three orders of magnitude in the past decade, is equal to [15]

$$\begin{aligned}\Delta\nu_{\rm IS} &= [E(2S) - E(1S)]_{\rm D} - [E(2S) - E(1S)]_{\rm H} \\ &= 670994334.64(15)\ \text{kHz},\end{aligned} \quad (1)$$

and the 1*S*–2*S* interval in hydrogen was measured with a record accuracy of several hundredths of kilohertz [16]:

$$\begin{aligned}\Delta\nu_{1S-2S}({\rm H}) &= 2466061413187103(46)\ \text{Hz}, \\ \delta &= 1.9 \times 10^{-14}.\end{aligned} \quad (2)$$

Experimental investigations of intervals (1) and (2) in muonic hydrogen are yet under preliminary preparation.

Various contributions to the energy levels of muonic atoms were theoretically evaluated many years ago in [17–19] (see also other references in review [1]). In recent years, various corrections in the energy spectrum of muonic hydrogen were primarily calculated for the 2*P*–2*S* Lamb shift and the hyperfine structure of the *S* levels [12, 20–23]. In those works, a particle interaction operator was constructed, which provided α^5- and α^6-order corrections for the 2*P*–2*S* interval and the hyperfine splitting of the 1*S* and 2*S* levels (α is the fine structure constant). At present, it is necessary to theoretically analyze corrections of orders α^3, α^4, and α^5 in the Lamb shift of the 1*S* and 2*S* levels of muonic hydrogen and muonic deuteron, in the μp–μd isotope shift for the 1*S*–2*S* transition, which remains unknown to date, and in the fine structure interval $E(1S) - 8E(2S)$. Such calculations may promote more active experimental investigations of fine structure intervals (1) and (2) in

muonic hydrogen and, as a result, more accurate determination of fundamental physical constants such as the charge radii of the proton and deuteron and the muon mass.

In this work, numerical results are obtained for contributions of the α^3, α^4, and α^5 orders in the 1*S*–2*S* splitting and in the μp–μd isotope shift for this splitting. Numerical values of certain contributions are calculated by known analytical formulas. Most contributions for the 1*S* and 2*S* energy levels in muonic hydrogen and the isotope shift are obtained for the first time in the integral form, which is used for numerical estimates. The dependence of the overwhelming majority of corrections on the principal quantum number is nontrivial; i.e., it is not reduced to a factor of $1/n^3$. The cause of such a dependence is associated with the characteristic photon momenta and will be discussed below. The aim of this work is to calculate quantum-mechanical corrections in the muonic hydrogen–muonic deuteron isotope shift for the 1*S*–2*S* transition and in the fine structure interval $E(1S) - 8E(2S)$, as well as to obtain the numerical values of these quantities with an accuracy of 10^{-9}. These values can be considered as reliable checkpoints both for the realization of corresponding experiments and for the extraction of more accurate values for the charge radii of the proton and deuteron and the muon mass from these experimental data.

The fine structure of the energy spectrum of hydrogen-like atoms has long been studied using various methods [1, 14, 24]. With the accuracy to the $(Z\alpha)^4$ terms, the energy levels of the *S* states of a hydrogen-like atom consisting of the particles with masses m_1 and m_2 are given by the expression

$$E_n = m_1 + m_2 - \frac{\mu(Z\alpha)^2}{2n^2} - \frac{\mu(Z\alpha)^4}{2n^3}\left[1 - \frac{3}{4n} + \frac{\mu^2}{4m_1m_2n}\right]$$

$$= \begin{cases} \mu p(1S)\text{: } 1043927826470.3586 \text{ meV}; \\ \mu p(2S)\text{: } 1043929722866.0601 \text{ meV}, \\ \mu d(1S)\text{: } 1981268455762.7537 \text{ meV}; \\ \mu d(2S)\text{: } 1981270453188.8081 \text{ meV}. \end{cases} \quad (3)$$

Although the relative theoretical error in Eq. (3), which is attributed to uncertainties in the fine structure constant α and particle masses, is on the order of 10^{-7}, we present the numerical values in Eq. (3) with an accuracy of 0.0001 meV, which is important for analyzing various intervals of the fine structure of the energy spectrum. We used the following values of the fundamental physical constants [3]:

$$\alpha^{-1} = 137.03599976(50),$$
$$m_\mu = 0.105658357(5) \text{ GeV},$$
$$m_p = 0.938271998(38) \text{ GeV},$$
$$m_d = 1.875612762(75) \text{ GeV}.$$

The contribution of Eq. (3) to the μp–μd isotope shift for the 1*S*–2*S* transition is decisive (see table). At the same time, a number of important effects of both electromagnetic and strong interactions should be consistently taken into account in order to obtain the isotope shift with the accuracy to α^5 terms inclusively.

2. EFFECTS OF ONE- AND TWO-LOOP VACUUM POLARIZATION IN ONE-PHOTON INTERACTION

Our calculations of the energy spectra of hydrogen-like atoms are performed using the quasipotential method, where a bound state of two particles is described by the Schrödinger-type equation [25, 26]

$$[G^f]^{-1}\psi_M \equiv \left(\frac{b^2}{2\mu_R} - \frac{\mathbf{p}^2}{2\mu_R}\right)\psi_M(\mathbf{p}) = \int \frac{d\mathbf{q}}{(2\pi)^3} V(\mathbf{p}, \mathbf{q}, M)\psi_M(\mathbf{q}), \quad (4)$$

where

$$b^2 = E_1^2 - m_1^2 = E_2^2 - m_2^2,$$

$\mu_R = E_1E_2/M$ is the relativistic reduced mass, and $M = E_1 + E_2$ is the bound-state mass. The quasipotential of Eq. (5) is constructed in QED perturbation theory by means of off-shell two-particle scattering amplitude *T* projected onto the positive frequency states at zero relative energies of the particles:

$$V = V^{(1)} + V^{(2)} + V^{(3)} + \ldots, \quad T = T^{(1)} + T^{(2)} + T^{(3)} + \ldots, \quad (5)$$

$$V^{(1)} = T^{(1)}, \quad V^{(2)} = T^{(2)} - T^{(1)}G^fT^{(1)}, \ldots. \quad (6)$$

Since the muon mass is larger than the electron mass, the Bohr radius of μp is smaller than the Bohr radius of the hydrogen atom. As a result, the Bohr radius of μp and the electron Compton wavelength are of the same order of magnitude:

$$\frac{\hbar^2}{\mu e^2}\Big/\frac{\hbar}{m_e c} = 0.737384,$$

where m_e is the electron mass and μ is the reduced mass of two particles in the μp atom. For this reason, vacuum polarization effects in the energy spectrum of muonic hydrogen increase significantly [27]. Figure 1 shows one- and two-loop vacuum-polarization effects in the one-photon interaction.

In order to evaluate the contribution from the diagram shown in Fig. 1a (electron vacuum polarization) to

Corrections of the α^3, α^4, and α^5 orders in the Lamb shift of the $1S$ and $2S$ energy levels of muonic hydrogen and muonic deuterium and in the isotope shift ΔE_{IS}

Contribution to the atomic energy	μp, meV		μd, meV		ΔE_{IS}	Formula, Ref.
	$1S$	$2S$	$1S$	$2S$		
1	2	3	4	5	6	7
Fine-structure formula $E_n = m_1 + m_2 - \frac{\mu(Z\alpha)^2}{2n^2} - \frac{\mu(Z\alpha)^4}{2n^3} \times \left[1 - \frac{3}{4n} + \frac{\mu^2}{4nm_1m_2}\right]$	1043927826470.3586	1043929722866.0601	1981268455762.7537	1981270453188.8081	101030.3530	(3) [1]
$\alpha(Z\alpha)^2$-order contribution of one-loop vacuum polarization to 1γ interaction	–1898.8379	–219.5849	–2129.2820	–245.3205	204.7085	(5) (6) (13)
$\alpha(Z\alpha)^4$-order Wichmann–Kroll contribution	0.0114	0.0012	0.0126	0.0014	–0.0010	[28]
$\alpha^2(Z\alpha)^2$-order contribution of two-loop vacuum polarization (VP–VP) to 1γ interaction	–1.8816	–0.2426	–2.1871	–0.2811	0.2616	(15) (16)
$\alpha^2(Z\alpha)^2$-order contribution of two-loop vacuum polarization (2-loop) to 1γ interaction	–12.6144	–1.4112	–14.0141	–1.5606	1.2476	(19) (20)
$\alpha^3(Z\alpha)^2$-order contribution of three-loop vacuum polarization (VP–VP–VP) to 1γ interaction	–0.0029	–0.0003	–0.0034	–0.0004	0.0004	[30] (23) (24)
$\alpha^3(Z\alpha)^2$-order contribution of three-loop vacuum polarization (VP–2-loop) to 1γ interaction	–0.0223	–0.0030	–0.0251	–0.0035	0.0023	[30] (25) (26)
$\alpha^3(Z\alpha)^2$-order contribution of three-loop vacuum polarization ($\Pi^{(p6)}$) to 1γ interaction	–0.0340	–0.0045	–0.0380	–0.0050	0.0035	[30]
$\alpha^3(Z\alpha)^2$-order contribution of relativistic effects and vacuum polarization in first-order perturbation theory	0.1962	0.0249	0.2515	0.0322	–0.0480	[19] (29)
$\alpha^3(Z\alpha)^2$-order contribution of relativistic effects and vacuum polarization in second-order perturbation theory	–0.2644	–0.0559	–0.3194	–0.0696	0.0413	[19] (39)

Table 1. (Contd.)

1	2	3	4	5	6	7
$\alpha^2(Z\alpha)^2$-order contribution of two-loop vacuum polarization in second-order perturbation theory	–2.0343	–0.1532	–2.3675	–0.1750	0.3114	(41) (42)
$\alpha^3(Z\alpha)^2$-order contribution of three-loop vacuum polarization (VP–VP, VP) in second-order perturbation theory	–0.0061	–0.0002	–0.0073	–0.0005	0.0009	(43) (45)
$\alpha^3(Z\alpha)^2$-order contribution of three-loop vacuum polarization (2-loop VP, VP) in second-order perturbation theory	–0.0059	–0.0016	–0.0069	–0.0021	0.0005	(46) (47)
$\alpha(Z\alpha)^4$-order muon self-energy contribution and muon vacuum polarization contribution	5.1180	0.6543	0.9395	0.7594	–0.7164	[1] [19]
$\alpha(Z\alpha)^5$-order radiative corrections	0.0355	0.0044	0.0414	0.0052	–0.0051	[1]
Radiative corrections and $\alpha^2(Z\alpha)^4$-order vacuum polarization	0.0178	0.0025	0.0209	0.0029	–0.0027	[1] [19]
$(Z\alpha)^5$-order recoil correction	0.3009	0.0428	0.1781	0.0253	0.1053	[1]
$(Z\alpha)^4$-order nuclear structure contribution	38.5711	4.8214	213.4218	26.6825	–152.6597	(48) [1, 23]
$(Z\alpha)^5$-order nuclear structure contribution	–0.1464	–0.0183	–2.9384	–0.3674	2.4429	(51) [19, 23]
$\alpha(Z\alpha)^4$-order nuclear structure contribution and vacuum polarization	0.2127	0.0274	1.4155	0.1824	–1.0478	(52) (53)
$\alpha(Z\alpha)^4$-order nuclear structure contribution and vacuum polarization in second-order perturbation theory	0.1327	0.0135	0.8913	0.0898	–0.6823	(54) (55)
$(Z\alpha)^5$-order nuclear polarizability contribution	–0.1291	–0.0161	92.0511	11.5064	–80.6577	[38] [36, 39]
$\alpha(Z\alpha)^4$-order hadron vacuum polarization contribution	–0.0864	–0.0108	–0.1010	–0.0126	0.0128	[40] [41]
Resulting contribution	1043927824598.8893	1043929722650.1499	1981268453925.6873	1981270452980.2974	101003.3495	

the particle interaction operator, it is necessary to make the following change in the photon propagator [27]:

$$\frac{1}{k^2} \longrightarrow \frac{\alpha}{3\pi}\int_1^\infty ds \frac{\sqrt{s^2-1}(2s^2+1)}{s^4(k^2+4m_e^2s^2)}. \quad (7)$$

For the hydrogen atom,

$$(-k^2) = \mathbf{k}^2 \sim \mu_e^2(Z\alpha)^2 \sim m_e^2(Z\alpha)^2,$$

where μ_e is the reduced mass of two particles in the hydrogen atom. In this case, neglecting the first term in the denominator on the right-hand side of Eq. (7), one obtains

$$-\frac{\alpha}{15\pi m_e^2}.$$

At the same time, when

$$\mathbf{k}^2 \sim \mu^2(Z\alpha)^2 \sim m_1^2(Z\alpha)^2,$$

where m_1 is the muon mass, as in muonic hydrogen, the parameters $\mu\alpha$ and m_e are of the same order of magnitude and the expansion in α in the denominator of Eq. (7) is invalid. In this case (for muonic hydrogen), the particle interaction operator should be constructed in the one-photon approximation, using exact expression (7). In what follows, we will take into account that electron vacuum polarization makes α^3-, α^4-, and α^5-order contributions to the energy spectrum of the S states.

Taking into account Eq. (7), the modification of the Coulomb potential

$$V^{\mathrm{C}}(\mathbf{k}) = -Ze^2/\mathbf{k}^2$$

due to vacuum polarization is given by the following expression in the momentum representation [27]:

$$V^{\mathrm{C}}_{\mathrm{VP}}(\mathbf{k}) = -4\pi Z\alpha\frac{\alpha}{\pi}\int_1^\infty \frac{\sqrt{\xi^2-1}}{3\xi^4}\frac{(2\xi^2+1)}{\mathbf{k}^2+4m_e^2\xi^2}d\xi. \quad (8)$$

The Fourier transform of Eq. (8) yields the corresponding operator in the coordinate representation:

$$V^{\mathrm{C}}_{\mathrm{VP}}(r) = \frac{\alpha}{3\pi}\int_1^\infty d\xi \frac{\sqrt{\xi^2-1}(2\xi^2+1)}{\xi^4} \times\left(-\frac{Z\alpha}{r}\exp(-2m_e\xi r)\right). \quad (9)$$

This expression makes it possible to obtain the α^3-order correction for electron vacuum polarization in the energy spectrum of the 1S and 2S states of the muonic hydrogen atom. Using the wavefunctions of the 1S and 2S states in the form

$$\psi_{100}(r) = \frac{W^{3/2}}{\sqrt{\pi}}e^{-Wr},$$
$$\psi_{200}(r) = \frac{W^{3/2}}{2\sqrt{2\pi}}e^{-Wr/2}\left(1-\frac{Wr}{2}\right), \quad (10)$$

where $W = \mu Z\alpha$, we represent this correction as

$$\Delta E_{1\gamma,\,\mathrm{VP}}(1S) = -\frac{\mu(Z\alpha)^2\alpha}{3\pi}\int_1^\infty \rho(\xi)d\xi\frac{1}{p_1^2(\xi)}, \quad (11)$$

where

$$p_1(\xi) = 1+\frac{m_e\xi}{W}, \quad \rho(\xi) = \frac{\sqrt{\xi^2-1}(2\xi^2+1)}{\xi^4},$$

and

$$\Delta E_{1\gamma,\,\mathrm{VP}}(2S) = -\frac{\mu(Z\alpha)^2\alpha}{6\pi} \times\int_1^\infty \rho(\xi)d\xi\left(\frac{1}{p_2^2(\xi)}-\frac{2}{p_2^3(\xi)}+\frac{3}{2p_2^4(\xi)}\right), \quad (12)$$

where

$$p_2(\xi) = 1+\frac{2m_e\xi}{W}.$$

Electron vacuum polarization effects are very sensitive to the structure of the bound state, because the characteristic momentum of particles in the muonic hydrogen atom is equal to $\mu(Z\alpha)$. For this reason, the contribution of the amplitudes under consideration with electron vacuum polarization is not reduced to the factor

$$|\psi^{\mathrm{C}}(0)|^2 \sim 1/n^3.$$

Its dependence on the principal quantum number is more complex and is given by Eqs. (11) and (12). The

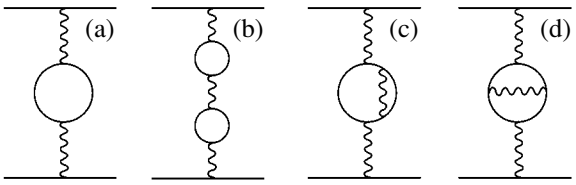


Fig. 1. One- and two-loop vacuum polarization effects in the one-photon interaction.

numerical values

$$\Delta E_{1\gamma,\,\mathrm{VP}} = \begin{cases} \mu p(1S)\colon -1898.8379\ \text{meV}; \\ \mu p(2S)\colon -219.5849\ \text{meV}, \\ \mu d(1S)\colon -2129.2820\ \text{meV}; \\ \mu d(2S)\colon -245.3205\ \text{meV}, \end{cases} \tag{13}$$

of the contribution of electron vacuum polarization for the $1S$ and $2S$ states of muonic hydrogen and muonic deuteron differ from each other due to the reduced mass of two particles. The contribution of muon vacuum polarization can be obtained from Eq. (8) by changing $m_e \longrightarrow m_1$. This α^5-correction in the energy spectrum of the muonic hydrogen-like atom is presented in the table along with the muon self-energy correction.

Let us consider the modification of the Coulomb potential due to the two-loop vacuum polarization (see Fig. 1). The contribution of the first diagram containing two sequential loops can be found by means of double change (7) in the photon propagator. In the coordinate representation, the corresponding particle interaction operator has the form

$$V^{\mathrm{C}}_{1\gamma,\,\mathrm{VP-VP}}(r) = \frac{\alpha^2}{9\pi^2}\int_1^\infty \rho(\xi)d\xi \int_1^\infty \rho(\eta)d\eta\left(-\frac{Z\alpha}{r}\right) \times \frac{1}{\xi^2-\eta^2}\left(\xi^2\exp(-2m_e\xi r) - \eta^2\exp(-2m_e\eta r)\right) \tag{14}$$

and makes the following contribution to the energy spectrum:

$$\Delta E_{1\gamma,\,\mathrm{VP-VP}}(1S) = -\frac{\mu\alpha^2(Z\alpha)^2}{9\pi^2} \times \int_1^\infty \rho(\xi)d\xi \int_1^\infty \rho(\eta)d\eta \frac{1}{\xi^2-\eta^2}\left(\frac{\xi^2}{p_1^2(\xi)} - \frac{\eta^2}{p_1^2(\eta)}\right) \tag{15}$$

$$= \begin{cases} \mu p\colon -1.8816\ \text{meV}, \\ \mu d\colon -2.1871\ \text{meV}, \end{cases}$$

$$\Delta E_{1\gamma,\,\mathrm{VP-VP}}(2S) = -\frac{\mu\alpha^2(Z\alpha)^2}{18\pi^2} \times \int_1^\infty \rho(\xi)d\xi \int_1^\infty \rho(\eta)d\eta \frac{1}{\xi^2-\eta^2} \times \left[\xi^2\left(\frac{1}{p_2^2(\xi)} - \frac{2}{p_2^3(\xi)} + \frac{3}{2p_2^4(\xi)}\right) - \eta^2\left(\frac{1}{p_2^2(\eta)} - \frac{2}{p_2^3(\eta)} + \frac{2}{2p_2^4(\eta)}\right)\right] \tag{16}$$

$$= \begin{cases} \mu p\colon -0.2426\ \text{meV}, \\ \mu d\colon -0.2811\ \text{meV}. \end{cases}$$

To calculate the contributions from the diagrams shown in Figs. 1c and 1d, which are determined by the second-order polarization operator, it is necessary to use the following substitution into the photon propagator [29]:

$$\frac{1}{k^2} \longrightarrow \left(\frac{\alpha}{\pi}\right)^2 \int_0^1 \frac{f(v)}{4m_e^2 + k^2(1-v^2)}dv$$

$$= \left(\frac{\alpha}{\pi}\right)^2 \frac{2}{3}\int_0^1 dv \frac{v}{4m_e^2 + k^2(1-v^2)}$$

$$\times\left\{(3-v^2)(1+v^2)\left[\mathrm{Li}_2\left(-\frac{1-v}{1+v}\right) + 2\mathrm{Li}_2\left(\frac{1-v}{1+v}\right) + \frac{3}{2}\ln\frac{1+v}{1-v}\ln\frac{1+v}{2} - \ln\frac{1+v}{1-v}\ln v\right]\right. \tag{17}$$

$$+\left[\frac{11}{16}(3-v^2)(1+v^2) + \frac{v^4}{4}\right]\ln\frac{1+v}{1-v}$$

$$\left. + \left[\frac{3}{2}v(3-v^2)\ln\frac{1-v^2}{4} - 2v(3-v^2)\ln v\right] + \frac{3}{8}v(5-3v^2)\right\}.$$

In this case, when calculating the numerical value of the contribution, it is convenient to use the coordinate representation and to reduce the particle interaction potential to the form

$$\Delta V^{\mathrm{C}}_{1\gamma,\,2\text{-loop VP}} = -\frac{2}{3}\frac{Z\alpha}{r} \times \left(\frac{\alpha}{\pi}\right)^2 \int_0^1 \frac{f(v)dv}{1-v^2}\exp\left(-\frac{2m_e r}{\sqrt{1-v^2}}\right). \tag{18}$$

This potential provides the following corrections in the Lamb shift of the S levels of muonic hydrogen and muonic deuteron:

$$\Delta E_{1\gamma,\,2\text{-loop VP}}(1S) = -\frac{2}{3\pi^2}\mu\alpha^2(Z\alpha)^2\frac{W^2}{m_e^2}$$

$$\times\int_0^1\frac{f(v)dv}{\left(1+\frac{W\sqrt{1-v^2}}{m_e}\right)^2} \tag{19}$$

$$=\begin{cases}\mu p\text{: }-12.6144\text{ meV},\\ \mu d\text{: }-14.0141\text{ meV},\end{cases}$$

$$\Delta E_{1\gamma,\,2\text{-loop VP}}(2S)=-\frac{1}{12\pi^2}\mu\alpha^2(Z\alpha)^2\frac{W^2}{m_e^2}$$

$$\times\int_0^1\frac{f(v)dv}{\left(1+\frac{W\sqrt{1-v^2}}{2m_e}\right)^2}\left[1-\frac{2}{p_3(v)}+\frac{3}{2p_3^2(v)}\right] \tag{20}$$

$$=\begin{cases}\mu p\text{: }-1.4112\text{ meV},\\ \mu d\text{: }-1.5606\text{ meV},\end{cases}$$

where

$$p_3(v)=\frac{1}{1+\frac{2m_e}{W\sqrt{1-v^2}}}.$$

Since we numerically evaluate the contribution to the energy spectrum, the corresponding results are presented with an accuracy of 0.0001 meV.

3. THREE-LOOP VACUUM POLARIZATION IN ONE-PHOTON INTERACTION

The three-loop vacuum polarization amplitudes in one-photon interaction also make α^5-order contribution (see Figs. 2a and 2b). The diagram shown in Fig. 2a makes the following contribution to the potential:

$$V^{\mathrm{C}}_{\mathrm{VP-VP-VP}}(r)=-\frac{Z\alpha}{r}\frac{\alpha^3}{(3\pi)^3}\int_1^\infty\rho(\xi)d\xi\times\int_1^\infty\rho(\eta)d\eta\int_1^\infty\rho(\zeta)d\zeta$$

$$\times\left[\exp(-2m_e\zeta r)\frac{\zeta^4}{(\xi^2-\zeta^2)(\eta^2-\zeta^2)}+\exp(-2m_e\xi r)\frac{\xi^4}{(\zeta^2-\xi^2)(\eta^2-\xi^2)}+\exp(-2m_e\eta r)\frac{\eta^4}{(\xi^2-\eta^2)(\zeta^2-\eta^2)}\right]. \tag{21}$$

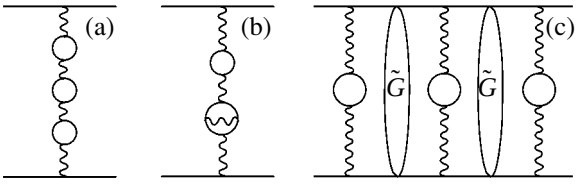


Fig. 2. Three-loop vacuum polarization effects in the (a, b) one-photon interaction and (c) third-order perturbation theory.

Using Eqs. (7) and (17), the contribution of the diagram shown in Fig. 2b to the potential can be represented in the integral form

$$V^{\mathrm{C}}_{\mathrm{VP-2\text{-}loop\ VP}}=-\frac{4\mu\alpha^3(Z\alpha)}{9\pi^3}$$

$$\times\int_1^\infty\rho(\xi)d\xi\int_1^\infty\frac{f(\eta)}{\eta}d\eta\frac{1}{r} \tag{22}$$

$$\times\left[\exp(-2m_e\eta r)\frac{\eta^2}{\eta^2-\xi^2}\exp(-2m_e\xi r)\frac{\xi^2}{\eta^2-\xi^2}\right].$$

The corrections to the energy spectrum of the μp and μd atoms that correspond to these interactions are

$$\Delta E_{\mathrm{VP-VP-VP}}(1S)=-\frac{\mu\alpha^3(Z\alpha)^2}{27\pi^3}$$

$$\times\int_1^\infty\rho(\xi)d\xi\int_1^\infty\rho(\eta)d\eta\int_1^\infty\rho(\zeta)d\zeta$$

$$\times\left[\frac{\xi^4}{(\xi^2-\eta^2)(\xi^2-\zeta^2)p_1^2(\xi)}+\frac{\eta^4}{(\eta^2-\xi^2)(\eta^2-\zeta^2)p_1^2(\eta)}+\frac{\zeta^4}{(\zeta^2-\xi^2)(\zeta^2-\eta^2)p_1^2(\zeta)}\right] \tag{23}$$

$$=\begin{cases}\mu p\text{: }-0.0029\text{ meV},\\ \mu d\text{: }-0.0034\text{ meV},\end{cases}$$

$$\Delta E_{\mathrm{VP-VP-VP}}(2S)=-\frac{\mu\alpha^3(Z\alpha)^2}{54\pi^3}$$

$$\times\int_1^\infty\rho(\xi)d\xi\int_1^\infty\rho(\eta)d\eta\int_1^\infty\rho(\zeta)d\zeta$$

$$\times\left\{\frac{\xi^4}{(\xi^2-\eta^2)(\xi^2-\zeta^2)}\left[\frac{1}{p_2^2(\xi)}-\frac{1}{p_2^3(\xi)}\right]\right.$$
$$+\frac{\eta^4}{(\eta^2-\xi^2)(\eta^2-\zeta^2)}\left[\frac{1}{p_2^2(\eta)}-\frac{1}{p_2^3(\eta)}\right]$$
$$\left.+\frac{\zeta^4}{(\zeta^2-\xi^2)(\zeta^2-\eta^2)}\left[\frac{1}{p_2^2(\zeta)}-\frac{1}{p_2^3(\zeta)}\right]\right\} \quad (24)$$
$$=\begin{cases}\mu p\text{: }-0.0003\text{ meV},\\ \mu d\text{: }-0.0004\text{ meV},\end{cases}$$

$$\Delta E_{\text{VP–2–loop VP}}(1S) = -\frac{4\mu\alpha^3(Z\alpha)^2}{9\pi^3}\int_1^\infty \rho(\xi)d\xi$$
$$\times\int_1^\infty \frac{f(\eta)}{\eta}\frac{d\eta}{\eta^2-\xi^2}\left[\frac{\eta^2}{p_1^2(\eta)}-\frac{\xi^2}{p_1^2(\xi)}\right] \quad (25)$$
$$=\begin{cases}\mu p\text{: }-0.0223\text{ meV},\\ \mu d\text{: }-0.0251\text{ meV},\end{cases}$$

$$\Delta E_{\text{VP–2–loop VP}}(2S) = -\frac{2\mu\alpha^3(Z\alpha)^2}{9\pi^3}$$
$$\times\int_1^\infty \rho(\xi)d\xi\int_1^\infty \frac{f(\eta)}{\eta}\frac{d\eta}{\eta^2-\xi^2}$$
$$\times\left\{\eta^2\left[\frac{1}{p_2^2(\eta)}-\frac{1}{p_2^3(\eta)}\right]-\xi^2\left[\frac{1}{p_2^2(\xi)}-\frac{1}{p_2^3(\xi)}\right]\right\} \quad (26)$$
$$=\begin{cases}\mu p\text{: }-0.0030\text{ meV},\\ \mu d\text{: }-0.0035\text{ meV}.\end{cases}$$

In the function $f(v)$ specified by Eq. (17), it is necessary to make the substitution

$$v = \frac{\sqrt{\eta^2-1}}{\eta}.$$

There are additional diagrams that express three-loop corrections in the polarization operator. They were first calculated for the $2P$–$2S$ Lamb shift by Kinoshita and Nio [30, 31]. The largest contribution to the energy spectrum comes from the sixth-order vacuum polarization diagrams with one-electron loop ($\Pi^{(p6)}$ contributions [30]). The estimate of their contribution to the Lamb shift of the $1S$ and $2S$ energy levels is presented in the table.

4. VACUUM POLARIZATION EFFECTS WITH THE INCLUSION OF RELATIVISTIC CORRECTIONS

Calculation of the energy spectrum of the S states of muonic hydrogen with an accuracy of α^5 requires the construction of the quasipotential using Eqs. (5) and (6) and taking into account relativistic effects of this order (Breit Hamiltonian ΔV_B). Taking into account electron vacuum polarization, the Breit Hamiltonian ΔV_B^{VP} was obtained by Pachucki [19]. Omitting the spin-dependent terms in the interaction operator, we represent these Hamiltonians in the form

$$\Delta V_B = -\mathbf{p}^4\left(\frac{1}{8m_1^3}+\frac{1}{8m_2^3}\right)$$
$$+\frac{\pi Z\alpha}{2}\left(\frac{1}{m_1^2}+\frac{1}{m_2^2}\right)\delta(\mathbf{r}) \quad (27)$$
$$-\frac{Z\alpha}{2m_1m_2r}\left(\mathbf{p}^2+\frac{r_ir_jp_ip_j}{r^2}\right),$$

$$\Delta V_B^{VP} = \frac{\alpha}{3\pi}\int_1^\infty \rho(\xi)d\xi\left\{\frac{Z\alpha}{2}\left(\frac{1}{m_1^2}+\frac{1}{m_2^2}\right)\right.$$
$$\times\left[\pi\delta(\mathbf{r})-\frac{m_e^2\xi^2}{r}\exp(2-m_e\xi r)\right]$$
$$-\frac{Z\alpha m_e^2\xi^2}{m_1m_2r}\exp(-2m_e\xi r)(1-m_e\xi r) \quad (28)$$
$$-\frac{Z\alpha}{2m_1m_2}p_i\frac{\exp(-2m_e\xi r)}{r}$$
$$\left.\times\left[\delta_{ij}+\frac{r_ir_j}{r^2}(1+2m_e\xi r)\right]p_j.\right.$$

In first-order perturbation theory, potential (28) makes the following contribution to the energy spectrum after averaging over Coulomb wavefunctions (10):

$$\Delta E_{1,B}^{\text{rel, VP}} = \langle\psi_{1S,2S}|\Delta V_B^{VP}|\psi_{1S,2S}\rangle$$
$$=\begin{cases}\mu p(1S)\text{: }0.1962\text{ meV};\ \mu p(2S)\text{: }0.0249\text{ meV},\\ \mu d(1S)\text{: }0.2515\text{ meV};\ \mu d(2S)\text{: }0.0322\text{ meV}.\end{cases} \quad (29)$$

These corrections are of the $\alpha(Z\alpha)^4$ order. The second-order perturbation corrections in the energy spectrum of the hydrogen-like system are determined by the

reduced Coulomb Green function [32], whose partial expansion has the form

$$\tilde{G}_n(\mathbf{r}, \mathbf{r}') = \sum_{l,m} \tilde{g}_{nl}(r, r') Y_{lm}(\mathbf{n}) Y^*_{lm}(\mathbf{n}'). \quad (30)$$

The radial function $\tilde{g}_{nl}(r, r')$ was obtained in [32] in the form of the Sturm expansion in Laguerre polynomials. For the 1*S* and 2*S* states, this function has the form

$$\tilde{g}_{10}(r, r') = -4\mu^2 Z\alpha \left(\sum_{m=2}^{\infty} \frac{L^1_{m-1}(x) L^1_{m-1}(x')}{m(m-1)} + \frac{5}{2} - \frac{x}{2} - \frac{x'}{2} \right) \exp\left(-\frac{x+x'}{2}\right), \quad (31)$$

$$\tilde{g}_{20}(r, r') = -2\mu^2 Z\alpha \left[\sum_{m=1, m\neq 2}^{\infty} \frac{L^1_{m-1}(x) L^1_{m-1}(x')}{m(m-2)} + \left(\frac{5}{2} + x\frac{\partial}{\partial x} + x'\frac{\partial}{\partial x'} \right) L^1_1(x) L^1_1(x') \right] \times \exp\left(-\frac{x+x'}{2}\right), \quad (32)$$

where $x = \mu Z\alpha r$ and

$$L^m_n(x) = \frac{e^x x^{-m}}{n!} \left(\frac{d}{dx}\right)^n (e^{-x} x^{n+m}) \quad (33)$$

are the Laguerre polynomials. Certain terms of the quasipotential include $\delta(\mathbf{r})$ and, therefore, it is necessary to know $\tilde{G}_n(\mathbf{r}, 0)$. The expression for the reduced Coulomb Green function in this case was derived in [33] using the Hostler representation for the Coulomb Green function as a result of the subtraction of the pole term in the form

$$\tilde{G}_{1S}(\mathbf{r}, 0) = \frac{Z\alpha\mu^2}{4\pi} \frac{2e^{-x/2}}{x} \times [2x(\ln x + C) + x^2 - 5x - 2], \quad (34)$$

$$\tilde{G}_{2S}(\mathbf{r}, 0) = -\frac{Z\alpha\mu^2}{4\pi} \frac{e^{-x/2}}{2x} \times [4x(x-2)(\ln x + C) + x^3 - 13x^2 + 6x + 4], \quad (35)$$

where $C = 0.5772\ldots$ is the Euler constant. In second-order perturbation theory, one-loop vacuum polarization makes the contribution

$$\Delta E_{n\,\mathrm{SOPT}} = 2 \sum_{m=1, m\neq n}^{\infty} \frac{\langle \psi^C_n | \Delta V_B | \psi^C_m \rangle \langle \psi^C_m | V^C_{\mathrm{VP}} | \psi^C_n \rangle}{E^C_n - E^C_m}. \quad (36)$$

Calculating the matrix element of the operator $\mathbf{p}^4$, we first use the substitution

$$\mathbf{p}^4/4\mu^2 = (H_0 + Z\alpha/r)(H_0 + Z\alpha/r),$$

$$H_0 = \mathbf{p}^2/2\mu - Z\alpha/r,$$

and then the algebraic transformations

$$\langle \psi^C_n | \frac{Z\alpha}{r} H_0 \sum_{m, m\neq n} \frac{|\psi^C_m\rangle\langle\psi^C_m|}{E^C_n - E^C_m} V^C_{\mathrm{VP}} | \psi^C_n \rangle = -\langle \psi^C_n | \frac{Z\alpha}{r} (I - |\psi^C_n\rangle\langle\psi^C_n|) V^C_{\mathrm{VP}} | \psi^C_n \rangle + E^C_n \langle \psi^C_n | \frac{Z\alpha}{r} \sum_{m, m\neq n} \frac{|\psi^C_m\rangle\langle\psi^C_m|}{E^C_n - E^C_m} V^C_{\mathrm{VP}} | \psi^C_n \rangle. \quad (37)$$

The sums arising after the integration with reduced Coulomb Green function (30) with respect to the coordinates r and r' are calculated taking into account that $1 < \xi < \infty$ and

$$\frac{1}{1 + W/m_e\xi} < 1.$$

The characteristic matrix element for the 1*S* state has the form

$$I = \int_0^\infty \left(1 - \frac{x}{2}\right) e^{-x} dx \int_0^\infty x' \exp(-x' p_1(\xi)) dx' \times \left[\frac{5}{2} - \frac{x}{2} - \frac{x'}{2} + \sum_m \frac{L^1_{m-1}(x) L^1_{m-1}(x')}{m(m-1)} \right]$$

$$= \frac{5}{4p_1^2(\xi)} - \frac{1}{2p_1^3(\xi)} + \frac{\ln p_1(\xi)}{p_1^2(\xi)}, \quad (38)$$

$$\sum_{m=2}^{\infty} \frac{[p_1(\xi) - 1]^{m-1}}{(m-1) p_1^{m+1}(\xi)} = \frac{1}{p_1^2(\xi)} \sum_{m=2}^{\infty} \frac{1}{(m-1) p_1^{m-1}(\xi)} = \frac{\ln p_1(\xi)}{p_1^2(\xi)}.$$

Omitting numerous other intermediate analytical expressions, we present the resulting numerical values

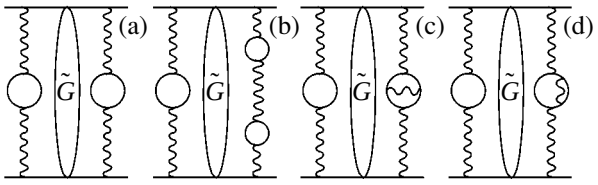


Fig. 3. Two- and three-loop vacuum polarization corrections in second-order perturbation theory.

of the corrections for the 1*S* and 2*S* levels in second-order perturbation theory (36):

$$\Delta E^{\mathrm{rel,\,VP}}_{2,\,\mathrm{B}} = \begin{cases} \mu p(1S)\text{: } -0.2644 \text{ meV};\ \ \mu p(2S)\text{: } -0.0559 \text{ meV}, \\ \mu d(1S)\text{: } -0.3194 \text{ meV};\ \ \mu d(2S)\text{: } -0.0696 \text{ meV}. \end{cases} \tag{39}$$

5. TWO- AND THREE-LOOP VACUUM POLARIZATIONS IN SECOND-ORDER PERTURBATION THEORY

Two-loop vacuum polarization makes the following contribution in second-order perturbation theory (see Fig. 3a):

$$\Delta E^{\mathrm{VP,\,VP}}_{\mathrm{SOPT}} = \langle \psi^{\mathrm{C}}_n | V^{\mathrm{C}}_{\mathrm{VP}} \tilde{G} V^{\mathrm{C}}_{\mathrm{VP}} | \psi^{\mathrm{C}}_n \rangle. \tag{40}$$

This matrix element can be calculated using Eqs. (9), (31), and (32). As a result, we arrive at the following $\alpha^2(Z\alpha)^2$-order corrections for the 1*S* and 2*S* energy levels:

$$\Delta E^{\mathrm{VP,\,VP}}_{\mathrm{SOPT}}(1S) = -\frac{\mu\alpha^2(Z\alpha)^2}{9\pi^2}\int_1^\infty \rho(\xi)d\xi \int_1^\infty \rho(\eta)d\eta \times \Bigg[\frac{5}{2}\frac{1}{p_1^2(\xi)p_1^2(\eta)} - \frac{1}{p_1^2(\xi)p_1^3(\eta)} - \frac{1}{p_1^3(\xi)p_1^2(\eta)} + \frac{[p_1(\xi)-1][p_1(\eta)-1]}{p_1^2(\xi)p_1^2(\eta)(p_1(\xi)+p_1(\eta)-1)} - \frac{1}{p_1^2(\xi)p_1^2(\eta)}\ln\frac{p_1(\xi)+p_1(\eta)-1}{p_1(\xi)+p_1(\eta)}\Bigg], \tag{41}$$

$$\Delta E^{\mathrm{VP,\,VP}}_{\mathrm{SOPT}}(2S) = -\frac{\mu\alpha^2(Z\alpha)^2}{9\pi^2} \times \int_1^\infty \rho(\xi)d\xi \int_1^\infty \rho(\eta)d\eta f(\xi,\eta). \tag{42}$$

Here,

$$\begin{aligned} f(\xi,\eta) &= -\frac{9}{2p_2^4(\xi)p_2^5(\eta)} + \frac{6}{p_2^3(\xi)p_2^5(\eta)} - \frac{3}{p_2^2(\xi)p_2^5(\eta)} \\ &- \frac{9}{2p_2^5(\xi)p_2^4(\eta)} + \frac{153}{8p_2^4(\xi)p_2^4(\eta)} - \frac{39}{2p_2^3(\xi)p_2^4(\eta)} \\ &+ \frac{33}{4p_2^2(\xi)p_2^4(\eta)} + \frac{6}{p_2^5(\xi)p_2^3(\eta)} - \frac{39}{2p_2^4(\xi)p_2^3(\eta)} \\ &+ \frac{17}{p_2^3(\xi)p_2^3(\eta)} - \frac{6}{p_2^2(\xi)p_2^3(\eta)} - \frac{3}{p_2^5(\xi)p_2^2(\eta)} \\ &+ \frac{33}{4p_2^4(\xi)p_2^2(\eta)} - \frac{6}{p_2^3(\xi)p_2^2(\eta)} + \frac{3}{2p_2^2(\xi)p_2^2(\eta)} \\ &- \frac{z^4}{a_2^2(\xi)a_2^2(\eta)(z-1)} - \frac{z^4(6-5z)}{2a_2^2(\xi)a_2^2(\eta)(z-1)^2p_2(\xi)} \\ &- \frac{z^4(6-5z)}{2a_2^2(\xi)a_2^2(\eta)(z-1)^2p_2(\eta)} + \frac{z^4(4-3z)}{2a_2^2(\xi)a_2^3(\eta)(z-1)^2} \\ &+ \frac{z^4(4-3z)}{2a_2^3(\xi)a_2^2(\eta)(z-1)^2} + \frac{z^5(-30+49z-21z^2)}{4a_2^3(\xi)a_2^3(\eta)(z-1)^3} \\ &- \frac{z^4(-18+27z-11z^2)}{4a_2^2(\xi)a_2^3(\eta)(z-1)^3p_2(\xi)} \\ &- \frac{z^4(-18+27z-11z^2)}{4a_2^3(\xi)a_2^2(\eta)(z-1)^3p_2(\eta)} \\ &+ \frac{z^4(-10+13z-5z^2)}{4a_2^3(\xi)a_2^3(\eta)(z-1)^3} + \ln(1-z) \\ &\times \Bigg[-\frac{2z^3}{a_2^2(\xi)a_2^2(\eta)} + \frac{3z^3}{a_2^2(\xi)a_2^2(\eta)p_2(\xi)} \\ &+ \frac{3z^3}{a_2^2(\xi)a_2^2(\eta)p_2(\eta)} - \frac{z^3}{a_2^2(\xi)a_2^3(\eta)} \\ &- \frac{z^3}{a_2^3(\xi)a_2^2(\eta)} - \frac{9z^4}{2a_2^3(\xi)a_2^3(\eta)} + \frac{3z^3}{2a_2^2(\xi)a_2^3(\eta)p_2(\xi)} \\ &+ \frac{3z^3}{2a_2^3(\xi)a_2^2(\eta)p_2(\eta)} - \frac{z^3}{2a_2^3(\xi)a_2^3(\eta)}\Bigg], \end{aligned}$$

where

$$a_2(\xi) = p_2(\xi) - 1, \quad z = \frac{a_2(\xi)a_2(\eta)}{p_2(\xi)p_2(\eta)}.$$

The contributions from three-loop vacuum polarization in second-order perturbation theory that are shown in Figs. 3b, 3c, and 3d are of the $\alpha^3(Z\alpha)^2$ order. To calculate them, it is necessary to use the expressions for one- and two-loop polarization operators. Integrating with respect to the coordinates of the wavefunctions, we represent these contributions in the form

$$\Delta E_{\mathrm{SOPT}}^{\mathrm{VP-VP,\,VP}}(1S) = -\frac{2\mu\alpha^3(Z\alpha)^2}{27\pi^3}$$
$$\times\int_1^\infty \rho(\xi)d\xi\int_1^\infty\frac{\rho(\eta)d\eta}{\xi^2-\eta^2} \tag{43}$$
$$\times\int_1^\infty\frac{\rho(\zeta)d\zeta}{p_1^2(\zeta)}\left[\frac{\xi^2 g(\xi,\zeta)}{p_1^2(\xi)} - \frac{\eta^2 g(\eta,\zeta)}{p_1^2(\eta)}\right],$$

where

$$g(\xi,\zeta) = \frac{5}{2} - \frac{1}{p_1(\xi)} - \frac{1}{p_1(\zeta)}$$
$$+\frac{[p_1(\xi)-1][p_1(\zeta)-1]}{p_1(\xi)+p_1(\zeta)-1} - \ln\frac{p_1(\xi)+p_1(\zeta)-1}{p_1(\xi)p_1(\zeta)}, \tag{44}$$

$$\Delta E_{\mathrm{SOPT}}^{\mathrm{VP-VP,\,VP}}(2S) = -\frac{2\mu\alpha^3(Z\alpha)^2}{27\pi^3}\int_1^\infty\rho(\xi)d\xi$$
$$\times\int_1^\infty\rho(\eta)d\eta\int_1^\infty\rho(\zeta)d\zeta\frac{\xi^2 f(\xi,\zeta)-\eta^2 f(\eta,\zeta)}{\xi^2-\eta^2}, \tag{45}$$

where the function $f(\xi, \zeta)$ is specified by Eq. (42);

$$\Delta E_{\mathrm{SOPT}}^{\text{2-loop VP, VP}}(1S) = -\frac{\mu\alpha^3(Z\alpha)^2}{9\pi^2}\int_0^1 f(v)dv$$
$$\times\left\{\frac{5}{2}\frac{1}{p_1^2(\xi)q_1^2(v)} - \frac{1}{p_1^3(\xi)q_1^2(v)} + \frac{1}{p_1^2(\xi)q_1^2(v)}\right.$$
$$\left.\times\left[\frac{p_1(\xi)q_1(v)}{p_1(\xi)+q_1(v)+1} - \ln\frac{p_1(\xi)+q_1(v)+1}{p_1(\xi)+q_1(v)}\right]\right\}, \tag{46}$$

where

$$q_1(v) = 1 + \frac{m_e}{W\sqrt{1-v^2}},$$

and

$$\Delta E_{\mathrm{SOPT}}^{\text{2-loop VP, VP}}(2S) = -\frac{4\mu\alpha^3(Z\alpha)^2}{9\pi^3}$$
$$\times\int_0^1\frac{f(v)dv}{1-v^2}\int_1^\infty\rho(\xi)d\xi f(\xi,v), \tag{47}$$

where

$$p_2(v) = 1 + \frac{2m_e}{W\sqrt{1-v^2}}, \quad a_2(v) = \frac{2m_e}{W\sqrt{1-v^2}}.$$

in the expression for the function $f(\xi, v)$. The contribution of the corrections given by Eqs. (41)–(47) to the isotope shift is equal to 0.3125 meV and individual contributions for the *S* levels are presented in the table. Analysis of the contribution of three-loop vacuum polarization in third-order perturbation theory in Fig. 2c shows that it is an order of magnitude smaller than the contribution from the diagram shown in Fig. 3c and thereby can be neglected.

6. EFFECTS OF THE STRUCTURE AND POLARIZABILITY OF A NUCLEUS AND VACUUM POLARIZATION

Strong interactions, which are associated with the distributions of the electric charge and magnetic moment of a nucleus, play an important role in the energy spectrum of muonic hydrogen. In the leading $(Z\alpha)^4$ order, the nuclear-structure effects are determined by the nuclear charge radius r_N, which is a differential characteristic of the electric-charge distribution. To calculate one-loop corrections, it is necessary to know the form of electromagnetic form factors of the nucleus. The contribution of nuclear structure effects both to the hyperfine structure of energy levels and to the Lamb shift was studied in [1, 19, 20, 23, 36]. The leading nuclear-structure correction of the $(Z\alpha)^4$ order in the energy spectrum of the *S* levels of muonic hydrogen has the form (see the diagram shown in Fig. 4a)

$$\Delta E_{\mathrm{str},\,(Z\alpha)^4}(nS) = \frac{2}{3n^3}\mu^3(Z\alpha)^4\langle r_p^2\rangle, \tag{48}$$

where r_p is the rms proton radius. The numerical values of this correction for the levels with $n = 1$ and 2 (at r_p = 0.891 fm [1]) that are presented in the table have a significant relative weight. For this reason, the reduction of the error in determining the proton charge radius is a

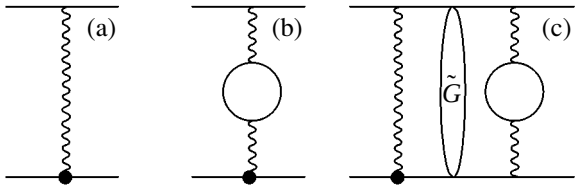


Fig. 4. Nuclear-structure and vacuum-polarization corrections.

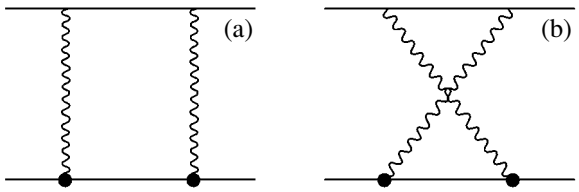


Fig. 5. Nuclear-structure corrections of the $(Z\alpha)^5$ order. The thick points in the diagrams are the nuclear vertex operators.

very urgent problem for obtaining a more accurate theoretical value for the Lamb shift of the S levels. In the case of deuterium, we use the value $r_d = 2.094$ fm for the deuteron charge radius [23].

Two-photon exchange amplitudes shown in Fig. 5 provide $(Z\alpha)^5$-order corrections to the nuclear structure. In this case, the two-photon interaction quasipotential can be evaluated using Eqs. (5) and (6) [19, 20] and the corresponding correction to the energy spectrum has the form of the one-dimensional integral

$$\Delta E_{\mathrm{str},(Z\alpha)^5}(nS) = -\frac{\mu^3(Z\alpha)^5}{\pi n^3}\int_0^\infty \frac{dk}{k}V(k), \qquad (49)$$

where

$$V(k) = \frac{2(F_1^2-1)}{m_1m_2} + \frac{8m_1[F_2(0)+4m_2^2F_1'(0)]}{m_2(m_1+m_2)k}$$

$$+\frac{k^2}{2m_1^3m_2^3}[2(F_1^2-1)(m_1^2+m_2^2)+4F_1F_2m_1^2+3F_2^2m_1^2]$$

$$+\frac{\sqrt{k^2+4m_1^2}}{2m_1^3m_2(m_1^2-m_2^2)k}$$

$$\times\left\{k^2[2(F_1^2-1)m_2^2+4F_1F_2m_1^2+3F_2^2m_1^2]\right. \qquad (50)$$

$$\left.-8m_1^4F_1F_2+\frac{16m_1^4m_2^2(F_1^2-1)}{k^2}\right\}$$

$$-\frac{\sqrt{k^2+4m_2^2}m_1}{2m_2^3(m_1^2-m_2^2)k}\left\{k^2[2(F_1^2-1)+4F_1F_2+3F_2^2]\right.$$

$$\left.-8m_2^2F_1F_2+\frac{16m_2^4(F_1^2-1)}{k^2}\right\}.$$

To perform numerical calculations based on Eq. (49), we use the Dirac F_1 and Pauli F_2 form factors in the parameterization obtained in [37]. Similar contributions for muonic deuteron were found in [23]. The α^5-order contribution associated with the nuclear structure also comes from electron vacuum polarization effects shown in Figs. 4b and 4c. The particle interaction operator corresponding to the amplitude in Fig. 4b is given by the expression

$$\Delta V_{\mathrm{str,\,VP}}(r) = \frac{2\alpha(Z\alpha)r_N^2}{9}\int_1^\infty \rho(\xi)d\xi$$
$$\times\left[\pi\delta(\mathbf{r})-\frac{m_e^2\xi^2}{r}\exp(-2m_e\xi r)\right], \qquad (51)$$

and the contributions to the energy spectrum of the $1S$ and $2S$ levels have the form

$$\Delta E_{\mathrm{str,\,VP}}(1S) = \frac{2}{9}\alpha(Z\alpha)^4\mu^3r_N^2\int_1^\infty\frac{\rho(\xi)d\xi}{p_1^2(\xi)}$$
$$\times\left(1+\frac{2m_e\xi}{W}\right) = \begin{cases}\mu p\text{: }0.1991\text{ meV},\\ \mu d\text{: }1.4155\text{ meV},\end{cases} \qquad (52)$$

and

$$\Delta E_{\mathrm{str,\,VP}}(2S) = \frac{1}{36}\alpha(Z\alpha)^4\mu^3r_N^2\int_1^\infty\frac{\rho(\xi)d\xi}{p_1^2(\xi)}$$
$$\times\left[\left(1+\frac{4m_e\xi}{W}\right)-\frac{4m_e^2\xi^2}{W^2}\left(-\frac{2}{p_2(\xi)}+\frac{3}{2p_2^2(\xi)}\right)\right] \qquad (53)$$
$$= \begin{cases}\mu p\text{: }0.0257\text{ meV},\\ \mu d\text{: }0.1824\text{ meV},\end{cases}$$

respectively. The contribution from electron vacuum polarization and nuclear structure in second-order perturbation theory (see the diagram shown in Fig. 4c) is

determined by the reduced Coulomb Green function $\tilde{G}_n(r, 0)$ given by Eqs. (34) and (35). In this case, the contributions to the shift of the S levels are equal to

$$\Delta E_{\text{str, VP; SOPT}}(1S) = -\frac{2}{9\pi}\alpha(Z\alpha)^4\mu^3 r_N^2 \times \int_1^\infty \frac{\rho(\xi)d\xi}{p_1^3(\xi)}[2 - 3p_1(\xi) - 2p_1^2(\xi) - 2p_1(\xi)\ln p_1(\xi)] \quad (54)$$

$$= \begin{cases} \mu p\text{: } 0.1242 \text{ meV}, \\ \mu d\text{: } 0.8913 \text{ meV}, \end{cases}$$

$$\Delta E_{\text{str, VP; SOPT}}(2S) = \frac{1}{36\pi}\alpha(Z\alpha)^4\mu^3 r_N^2 \int_1^\infty \frac{\rho(\xi)d\xi}{p_1^5(\xi)} \times [-12 + 23p_2(\xi) - 8p_2^2(\xi) - 4p_2^3(\xi) + 4p_2^4(\xi) + 4p_2(\xi)(3 - 4p_2(\xi) + 2p_2^2(\xi))\ln p_2(\xi)] \quad (55)$$

$$= \begin{cases} \mu p\text{: } 0.0126 \text{ meV}, \\ \mu d\text{: } 0.0898 \text{ meV}. \end{cases}$$

Two additional contributions of strong interaction to the Lamb shift are caused by nuclear polarizability and hadron vacuum polarization. The contribution of nuclear polarizability to the shift of the S levels and the hyperfine structure was obtained in [38] for muonic deuteron and in [20, 21] for muonic hydrogen. The $\alpha(Z\alpha)^4$-order contribution from hadron vacuum polarization to the shift of the S levels was studied in [40, 41] and is also presented in the table.

7. CONCLUSIONS

In this work, various QED effects, effects of the structure and polarizability of the proton, and hadron vacuum polarization have been calculated for the Lamb shift of the 1S and 2S energy levels in muonic hydrogen and muonic deuteron, as well as for the μp–μd isotope shift for the 1S–2S splitting. Corrections of the α^3, α^4, and α^5 orders, as well as certain α^6-order contributions enhanced by $\ln\alpha$, are evaluated. We take into account that the ratio $\mu\alpha/m_e$ is close to unit and focus on calculation of the effects of electron vacuum polarization. The numerical values of the contributions obtained in this work were presented in the table. The table also contains the Wichmann–Kroll correction [1, 28], muon self-energy correction, and $\alpha(Z\alpha)^4$-order contribution from muon vacuum polarization [1, 19], $\alpha(Z\alpha)^5$-order radiative corrections with a coefficient of about 10 [1], α^6-order contributions from muon radiative effects with the inclusion of electron vacuum polarization [1, 34], and $(Z\alpha)^5$-order recoil correction from two-photon exchange diagrams, which is known in the analytical form [35]. References to many works are given, where similar corrections were analyzed analytically or numerically, although numerical results for the 1S and 2S levels are absent in those works. For comparison of the results obtained in this work with calculations by other authors, [1] is often used, because it accumulates the recent results for the energy spectrums of simple atoms and contains detailed references to previous investigations.

Let us list a number of the features of the calculations.

(i) For muonic hydrogen, vacuum polarization effects are important and give rise to the modification of both the Breit particle interaction potential and the total interaction operator, which provides the α^5-order corrections to the energy spectrum.

(ii) In the leading $(Z\alpha)^4$ order and in the one-loop amplitudes, proton-structure effects in the energy spectrum of the S states are expressed in terms of the charge radius of a nucleus (proton and deuteron) and in terms of the nuclear electromagnetic form factor, respectively.

(iii) Contributions to the nuclear polarizability (second contribution of strong interaction) are estimated using relations obtained in [20, 21, 38]. Contributions from the structure and polarizability of a nucleus introduce the largest theoretical error to the 1S–2S interval, as well as to the isotope shift for this splitting.

The resulting numerical values of the energies of the 1S and 2S states in muonic hydrogen and muonic deuteron, as well as the isotopic shift for the 1S–2S transition, are presented in the table. They can be treated as a reliable estimate for future experiments concerning both the muonic hydrogen–muonic deuteron isotope shift and the 1S–2S large fine-structure interval in muonic hydrogen and muonic deuteron. The numerical values of the corrections were calculated with an accuracy of 0.0001 meV. For the 1S–2S transition, the theoretical error of the results is determined by uncertainties of the fundamental parameters (the fine structure constant and the masses of the proton, deuteron, and muon), which are on the order of 10^{-7}. The α^6-order QED corrections provide an error on the order of 10^{-8}. Uncertainties in the charge radii of the proton and deuteron make the largest contribution to the theoretical error. Their relative contribution is on the order of 10^{-6} ($r_p = 0.891$ fm and $r_d = 2.094$ fm are used for the charge radii of the proton and deuteron, respectively). Further improvements of theoretical results presented in the table are primarily associated with corrections to the structure and polarizability of a nucleus. In view of this circumstance, the fine structure interval $E(1S) - 8E(2S)$, where the effects of the structure and polarizability of

the nucleus are absent in the leading $(Z\alpha)^4$-order correction, is of particular interest for the comparison of the performed theoretical calculations with experimental data. The numerical values for this interval for muonic hydrogen and muonic deuteron are

$$\begin{aligned} &\mu p\colon E(1S) - 8E(2S) \\ &= -7307509956602.3099 \text{ meV}, \\ &\mu d\colon E(1S) - 8E(2S) \\ &= -13868895169916.6917 \text{ meV}. \end{aligned} \tag{56}$$

The relative value of that part of the theoretical error in Eq. (56) which is attributed to higher order corrections is very small (on the order of 10^{-15}).

The isotope shift in the muonic hydrogen–muonic deuteron system for the $1S$–$2S$ splitting is an important characteristic of the energy spectrums of these hydrogen-like atoms. The difference in the wavelengths emitted by the isotopes of one element appears due to the difference in the isotope masses, as well as to the difference in the distributions of the charge of the nuclei. At present, the proton and deuteron masses are known with a quite high accuracy, whereas the characteristics of the nuclear structure are determined less accurately. Using the calculation results, one can express the difference of charge radii between the deuteron and proton in terms of the isotope shift:

$$\begin{aligned} &\frac{r_d^2}{(1 + m_\mu/m_d)^3} - \frac{r_p^2}{(1 + m_\mu/m_p)^3} \\ &= \frac{12}{7m_\mu^3(Z\alpha)^4}(\Delta\tilde{E}_{\rm IS}^{\rm th} - \Delta E_{\rm IS}^{\rm exp}), \end{aligned} \tag{57}$$

where the theoretical value $\Delta\tilde{E}_{\rm IS}^{\rm th}$ does not include the $(Z\alpha)^4$-order correction to the nuclear structure. Thus, measurement of the μp–μd isotope shift would provide an additional test of QED and make it possible to obtain a more accurate value of the deuteron charge radius from Eq. (57) after the determination of the proton charge radius in the PSI experiment [4].

ACKNOWLEDGMENTS

I am grateful to D.D. Bakalov, T. Kinoshita, V.A. Matveev, M. Nio, V.G. Pal'chikov, K. Pakhutskiĭ, V.A. Rubakov, V.I. Savrin, R.N. Faustov, V.V. Fil'chenkov, I.B. Khriplovich, and M.I. Éĭdes for stimulating discussions on various aspects of the problem of the energy spectrum of the muonic hydrogen atom. This work was supported by the Russian Foundation for Basic Research (project no. 04-02-16085) and the program "Russian Universities–Fundamental Research" (project no. UR.02.01.367).

REFERENCES

1. M. I. Eides, H. Grotch, and V. A. Shelyuto, Phys. Rep. **342**, 62 (2001).
2. Yu. L. Sokolov, Usp. Fiz. Nauk **169**, 559 (1999) [Phys. Usp. **42**, 481 (1999)].
3. P. J. Mohr and B. N. Taylor, Rev. Mod. Phys. **72**, 351 (2000).
4. R. Pohl, F. Biraben, C. A. N. Conde, *et al.*, Hyperfine Interact. **127**, 161 (2000).
5. A. Adamczak, D. Bakalov, K. Bakalova, *et al.*, Hyperfine Interact. **136**, 1 (2001).
6. A. Dupays, A. Beswick, B. Lepetit, *et al.*, Phys. Rev. A **68**, 052503 (2003).
7. A. C. Zemach, Phys. Rev. **104**, 1771 (1956).
8. A. P. Martynenko and R. N. Faustov, Zh. Éksp. Teor. Fiz. **125**, 48 (2004) [JETP **98**, 39 (2004)]; hep-ph/0312116.
9. A. P. Martynenko, Phys. Rev. A **71**, 022506 (2005).
10. S. D. Drell and J. D. Sullivan, Phys. Rev. **154**, 1477 (1967).
11. J. Bernabeu and T. E. O. Ericson, Z. Phys. A **309**, 213 (1983).
12. E. V. Cherednikova, R. N. Faustov, and A. P. Martynenko, Nucl. Phys. A **703**, 365 (2002).
13. G. Breit, Rev. Mod. Phys. **30**, 507 (1958).
14. K. Pachucki, D. Leibfried, M. Weitz, *et al.*, J. Phys. B **29**, 177 (1996).
15. A. Huber, Th. Udem, B. Gross, *et al.*, Phys. Rev. Lett. **80**, 468 (1998).
16. M. Niering, R. Holzwarth, J. Reichert, *et al.*, Phys. Rev. Lett. **84**, 5496 (2000).
17. M. M. Sternheim, Phys. Rev. **138**, B430 (1965).
18. E. Borie and G. A. Rinker, Rev. Mod. Phys. **54**, 67 (1982).
19. K. Pachucki, Phys. Rev. A **53**, 2092 (1996).
20. A. P. Martynenko and R. N. Faustov, Yad. Fiz. **63**, 915 (2000) [Phys. At. Nucl. **63**, 845 (2000)].
21. K. Pachucki, Phys. Rev. A **60**, 3593 (1999).
22. A. Veitia and K. Pachucki, Phys. Rev. A **69**, 042501 (2004).
23. R. N. Faustov and A. P. Martynenko, Phys. Rev. A **67**, 052506 (2003).
24. E. E. Salpeter, Phys. Rev. **87**, 328 (1952); **89**, 92 (1953).
25. A. P. Martynenko and R. N. Faustov, Teor. Mat. Fiz. **64**, 179 (1985).
26. A. P. Martynenko and R. N. Faustov, Zh. Éksp. Teor. Fiz. **115**, 1221 (1999) [JETP **88**, 672 (1999)].
27. V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics*, 2nd ed. (Nauka, Moscow, 1980; Pergamon, Oxford, 1982).
28. E. H. Wichmann and N. M. Kroll, Phys. Rev. **101**, 843 (1956).
29. M. I. Eides, H. Grotch, and V. A. Shelyuto, Phys. Rev. D **65**, 013003 (2002).
30. T. Kinoshita and M. Nio, Phys. Rev. Lett. **82**, 3240 (1999).

31. T. Kinoshita and M. Nio, Phys. Rev. D **60**, 053008 (1999).

32. S. A. Zapryagaev, N. L. Manakov, and V. G. Pal'chikov, *Theory of Multi-charge Ions with One and Two Electrons* (Énergoatomizdat, Moscow, 1985) [in Russian].

33. V. G. Ivanov and S. G. Karshenboĭm, Zh. Éksp. Teor. Fiz. **109**, 1219 (1996) [JETP **82**, 656 (1996)].

34. R. Barbieri, M. Caffo, and E. Remiddi, Lett. Nuovo Cimento **7**, 60 (1973).

35. J. R. Sapirstein and D. R. Yennie, in *Quantum Electrodynamics*, Ed. by T. Kinoshita (World Sci., Singapore, 1990), p. 560.

36. J. L. Friar, J. Martorell, and D. W. L. Sprung, Phys. Rev. A **56**, 4579 (1997).

37. G. G. Simon, Ch. Schmitt, F. Borkowski, *et al.*, Nucl. Phys. A **333**, 381 (1980).

38. A. I. Mil'shteĭn, S. S. Petrosyan, and I. B. Khriplovich, Zh. Éksp. Teor. Fiz. **109**, 1146 (1996) [JETP **82**, 616 (1996)]; A. I. Mil'shteĭn and I. B. Khriplovich, Zh. Éksp. Teor. Fiz. **125**, 205 (2004) [JETP **98**, 181 (2004)].

39. R. N. Faustov and A. P. Martynenko, Mod. Phys. Lett. A **16**, 507 (2001).

40. J. L. Friar, J. Martorell, and D. W. L. Sprung, Phys. Rev. A **59**, 4061 (1999).

41. R. N. Faustov and A. P. Martynenko, Eur. Phys. J. C **6**, 1 (1999).

Translated by R. Tyapaev

Journal of Experimental and Theoretical Physics, Vol. 101, No. 6, 2005, pp. 1036–1052.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 6, 2005, pp. 1184–1200.

NUCLEI, PARTICLES, FIELDS, GRAVITATION, AND ASTROPHYSICS

Gravitating Global Monopoles in Extra Dimensions and the Braneworld Concept¶

K. A. Bronnikov[a,b] and B. E. Meierovich[c,*]

[a]*Center for Gravitation and Fundamental Metrology, Russian Research Institute for Metrological Service, Moscow, 117313 Russia*

[b]*Institute of Gravitation and Cosmology, Russian University of Peoples' Friendship, Moscow, 117198 Russia*

[c]*Kapitza Institute for Physical Problems, Moscow, 117334 Russia*

**e-mail: meierovich@yahoo.com; kb20@yandex.ru*

Received June 15, 2005

Abstract—Multidimensional configurations with a Minkowski external spacetime and a spherically symmetric global monopole in extra dimensions are discussed in the context of the braneworld concept. The monopole is formed with a hedgehoglike set of scalar fields ϕ^i with a symmetry-breaking potential V depending on the magnitude $\phi^2 = \phi^i\phi^i$. All possible kinds of globally regular configurations are singled out without specifying the shape of $V(\phi)$. These variants are governed by the maximum value ϕ_m of the scalar field, characterizing the energy scale of symmetry breaking. If $\phi_m < \phi_{cr}$ (where ϕ_{cr} is a critical value of ϕ related to the multidimensional Planck scale), the monopole reaches infinite radii, whereas in the "strong field regime," when $\phi_m \geq \phi_{cr}$, the monopole may end with a finite-radius cylinder or have two regular centers. The warp factors of monopoles with both infinite and finite radii may either exponentially grow or tend to finite constant values far from the center. All such configurations are shown to be able to trap test scalar matter, in striking contrast to RS2 type five-dimensional models. The monopole structures obtained analytically are also found numerically for the Mexican hat potential with an additional parameter acting as a cosmological constant. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

According to a presently popular idea, our observable Universe can be located on a four-dimensional surface, called the brane, embedded in a higher-dimensional manifold, called the bulk. This "braneworld" concept, suggested in the 1980s [1], is broadly discussed nowadays, mainly in connection with the recent developments in supersymmetric string/M-theories [2]. One reason why we do not see any extra dimensions is that the observed matter is confined to the brane and only gravity propagates in the bulk. There are numerous applications of the braneworld concept to particle physics, astrophysics, and cosmology, such as the hierarchy problem and the description of dark matter and dark energy [3].

Most of the studies are restricted to infinitely thin branes with deltalike localization of matter. A well-known example is Randall and Sundrum's second model (RS2) [4], in which a single Minkowski brane is embedded in a five-dimensional anti-de Sitter (AdS) bulk.

Thin branes can, however, only be treated as a rough approximation, because any fundamental underlying theory, be it quantum gravity or string or M-theory, must contain a fundamental length beyond which the classical spacetime description is impossible. It is therefore necessary to justify the infinitely thin brane approximation as a well-defined limit of a smooth structure, a thick brane, obtainable as a solution of coupled gravitational and matter field equations. Such a configuration is then required to be globally regular, stable, and properly concentrated around a three-dimensional surface that is meant to describe the observed spatial dimensions. Topological defects emerging in phase transitions with spontaneous symmetry breaking (SSB) are probably the best candidates for this role.

It should be mentioned that the evolution of the Universe, according to modern views, contained a sequence of phase transitions with SSB. A decisive step toward cosmological applications of the SSB concept was made in 1972 by Kirzhnits [5]. He assumed that, as in the case of solid substances, a symmetry of a field system, existing at sufficiently high temperatures, could be spontaneously broken as the temperature falls down. A necessary consequence of such phase transitions is the appearance of topological defects. The first quantitative analysis of the cosmological consequences of SSB was given by Zel'dovich, Kobzarev, and Okun' [6]. Later, the SSB phenomenon and various topological defects were widely used in inflationary

¶ The text was submitted by the authors in English.

Universe models and in attempts to explain the origin of the large-scale structure of the Universe (see, e.g., [7, 8]).

The properties of global topological defects are generally described with the aid of a multiplet of scalar fields playing the role of an order parameter. If a defect is to be interpreted as a braneworld, its structure is determined by the self-gravity of the scalar field system and may be described by a set of Einstein and scalar equations.

In this paper, we analyze the gravitational properties of candidate (thick) braneworlds with the four-dimensional Minkowski metric as global topological defects in extra dimensions. Our general formulation covers such particular cases as a brane (domain wall) in five-dimensional spacetime (one extra dimension), a global cosmic string with winding number $n = 1$ (two extra dimensions), and global monopoles (three or more extra dimensions). We restrict ourselves to Minkowski branes, because most of the existing problems are clearly seen even in these comparatively simple systems; on the other hand, in the majority of physical situations, the inner curvature of the brane itself is much smaller than the curvature related to brane formation, and, therefore, the main qualitative features of Minkowski branes should survive in curved branes.

Brane worlds as thick domain walls in a five-dimensional bulk have been discussed in many papers (see, e.g., [9] and references therein). Such systems were analyzed in a general form in [10, 11] without specifying the symmetry-breaking potential; it was shown, in particular, that all regular configurations should have an AdS asymptotic form. Therefore, all possible thick branes are merely regularized versions of the RS2 model, with all concomitant difficulties in matter-field confinement. Thus, it has been demonstrated [11] that a test scalar field has a divergent stress–energy tensor infinitely far from the brane, at the AdS horizon. The reason for that is the repulsive gravity of the RS2 and similar models: gravity repels matter from the brane and pushes it towards the AdS horizon. To overcome this difficulty, it is natural to try considering a greater number of extra dimensions. This was one of the reasons for us to consider higher-dimensional bulks.

We study the simplest possible realization of this idea, assuming a static, spherically symmetric configuration of the extra dimensions and a thick Minkowski brane as a concentration of the scalar field stress–energy tensor near the center. The possible trapping properties of gravity for test matter are then determined by the behavior of the so-called warp factor (the metric coefficient acting as a gravitational potential) far from the center, and we indeed find classes of regular solutions where gravity is attracting.

Some of our results repeat those obtained in [12, 13], which have discussed global and gauge ('t Hooft–Polyakov-type) monopoles in extra dimensions; a more detailed comparison is given in Section 7.

The paper is organized as follows. In Section 2, we formulate the problem, introduce spacetimes with global topological defects in the extra dimensions, write the equations and boundary conditions, and demonstrate a connection between the possibility of SSB and the properties of the potential at a regular center. In Section 3, we briefly discuss the trapping problem for RS2-type domain-wall models and show that they always have repulsive gravity and are unable to trap matter in the form of a test scalar field. Section 4 is devoted to a search for regular global monopole solutions in higher dimensions by analyzing their asymptotic properties far from the center. All regular configurations are classified by the behavior of the spherical radius r and by the properties of the potential. This leads to separation of "weak gravity" and "strong gravity" regimes, related to maximum values of the scalar field magnitude.

In the weak gravity regime, the spherical radius r tends to infinity along with the distance from the center. Such moderately curved configurations exist without any restrictions of fine-tuning type. If the scalar field magnitude exceeds some critical value, the radius r either tends to a finite value far from the center or returns to zero at a finite distance from the center, thus forming one more centers, which should also be regular. Some cases require fine tuning of the parameters of the potential, and, hence, one may believe that static configurations can only exist if the scalar and gravitational forces are somewhat mutually balanced.

In Section 5, we show that, in contrast to domain walls, global monopoles in different regimes do provide scalar field trapping on the brane. Section 6 is a brief description of numerical experiments with the Mexican hat potential admitting shifts up and down, equivalent to introducing a bulk cosmological constant. Their results confirm and illustrate the conclusions in Section 4. Section 7 summarizes the results.

2. PROBLEM SETTING

2.1. Geometry

We consider a $(D = d_0 + d_1 + 1)$-dimensional spacetime with the structure $\mathbb{M}^{d_0} \times \mathbb{R}_u \times \mathbb{S}^{d_1}$ and the metric

$$\begin{aligned} ds^2 &= e^{2\gamma(u)}\eta_{\mu\nu}dx^\mu dx^\nu \\ &- (e^{2\alpha(u)}du^2 + e^{2\beta(u)}d\Omega^2). \end{aligned} \tag{1}$$

Here,

$$\eta_{\mu\nu}dx^\mu dx^\nu = dt^2 - (d\mathbf{x})^2$$

is the Minkowski metric in the subspace $\mathbb{M}^{d_0}$,

$$\eta_{\mu\nu} = \mathrm{diag}(1, -1, \ldots, -1);$$

$d\Omega$ is a linear element on a d_1-dimensional unit sphere $\mathbb{S}^{d_1}$; α, β, and γ are functions of the radial coordinate u

with the definition domain $\mathbb{R}_u \subseteq \mathbb{R}$, to be specified later. The Riemann tensor has a diagonal form, and its nonzero components are

$$R^{\mu\nu}{}_{\rho\sigma} = -e^{-2\alpha}\gamma'^2\delta^{\mu\nu}_{\rho\sigma},$$

$$R^{ab}{}_{cd} = (e^{-2\beta} - e^{-2\alpha}\beta'^2)\delta^{ab}_{cd},$$

$$R^{u\mu}{}_{u\nu} = -\delta^\mu_\nu e^{-\gamma-\alpha}(e^{\gamma-\alpha}\gamma')', \quad (2)$$

$$R^{ua}{}_{ub} = -\delta^a_b e^{-\beta-\alpha}(e^{\beta-\alpha}\beta')',$$

$$R^{a\mu}{}_{b\nu} = -\delta^\mu_\nu\delta^a_b e^{-2\alpha}\gamma'\beta',$$

where

$$\delta^{\mu\nu}_{\rho\sigma} = \delta^\mu_\rho\delta^\nu_\sigma - \delta^\mu_\sigma\delta^\nu_\rho \quad (3)$$

and similarly for δ^{ab}_{cd}. Greek indices $\mu, \nu, \ldots$ correspond to the d_0-dimensional spacetime, and Latin indices $a, b, \ldots$ to d_1 angular coordinates on $\mathbb{S}^{d_1}$. We mostly bear in mind the usual dimension $d_0 = 4$, but keep d_0 arbitrary for generality.

A necessary condition of regularity is the finiteness of the Kretschmann scalar

$$\mathcal{K} = R^{AB}{}_{CD}R^{CD}{}_{AB}.$$

(Capital indices $A, B, \ldots$ correspond to all D coordinates.) In our case, $\mathcal{K}$ is a sum of squares of all nonzero $R^{AB}{}_{CD}$. Hence, in regular configurations, all components of Riemann tensor (2) are finite.

For the Ricci tensor, we have

$$R^\nu_\mu = -\delta^\nu_\mu e^{-2\alpha}[\gamma'' + \gamma'(-\alpha' + d_0\gamma' + d_1\beta)],$$

$$R^u_u = -e^{-2\alpha}[d_0(\gamma'' + \gamma'^2 - \alpha'\gamma')$$

$$+ d_1(\beta'' + \beta'^2 - \alpha'\beta')], \quad (4)$$

$$R^n_m = e^{-2\beta}(d_1 - 1)\delta^n_m$$

$$- \delta^n_m e^{-2\alpha}[\beta'' + \beta'(-\alpha' + d_0\gamma' + d_1\beta')].$$

2.2. Topological Defects

A global defect with a nonzero topological charge can be constructed as a multiplet of $d_1 + 1$ real scalar fields ϕ^k, in the same way as, e.g., in [14]. It comprises a hedgehog configuration in $\mathbb{R}_u \times \mathbb{S}^{d_1}$:

$$\phi^k = \phi(u)n^k(x^a),$$

where n^k is a unit vector in the $(d_1 + 1)$-dimensional Euclidean target space of the scalar fields:

$$n^k n^k = 1.$$

The total Lagrangian of the system is taken in the form

$$L = \frac{R}{2\kappa^2} + \frac{1}{2}g^{AB}\partial_A\phi^k\partial_B\phi^k - V(\phi), \quad (5)$$

where R is the D-dimensional scalar curvature, κ^2 is the D-dimensional gravitational constant, and V is a symmetry-breaking potential depending on $\phi^2(u) = \phi^a\phi^a$.

In the case where $d_1 = 0$, there is only one extra dimension. The topological defect is a flat domain wall. Combined with $d_0 = 4$, it is widely considered with reference to our Universe. Regular thick Minkowski branes supported by scalar fields with arbitrary potentials were analyzed in [10, 11] (see also Section 3 below).

The case where $d_1 = 1$ is a global cosmic string with the winding number $n = 1$. If $d_0 = 2$, it is a cosmic string in four dimensions, whose gravitational properties are reviewed in [15]. The case $d_0 = 4$ corresponds to a string in extra dimensions.

The case where $d_1 = 2$ and $d_0 = 1$ is a global monopole in our four-dimensional spacetime. We have analyzed it in detail in [16]. The case where $d_1 > 2$ and $d_0 = 1$ is its multidimensional generalization to static spherically symmetric spacetimes with d_1-dimensional rather than two-dimensional coordinate spheres [14]. It was shown that such a heavy multidimensional global monopole leads to a multidimensional cosmology where the symmetry-breaking potential at late times can mimic both dark matter and dark energy.

In the case where $d_0 = 4$ and $d_1 > 2$, we have a multidimensional global monopole entirely in the extra spacelike dimensions. Different models of this kind were studied in [12, 13, 17, 18]. In particular, such a monopole in extra dimensions was used in an attempt to explain the origin of inflation [17].

2.3. Field Equations

We use the Einstein equations in the form

$$R^B_A = -\kappa^2\tilde{T}^B_A, \quad \tilde{T}^B_A = T^B_A - \frac{\delta^B_A}{D-2}T^C_C,$$

where T^B_A is the stress-energy tensor of the scalar field

multiplet. For our hedgehog configuration,

$$\tilde{T}^{\nu}_{\mu} = -\frac{2V\delta^{\nu}_{\mu}}{D-2},$$

$$\tilde{T}^{u}_{u} = -\frac{2V}{D-2} - e^{-2\alpha}\phi'^{2},$$

$$\tilde{T}^{b}_{a} = -\frac{2V\delta^{b}_{a}}{D-2} - e^{-2\beta}\delta^{b}_{a}\phi'^{2}.$$

So far, we did not specify the radial coordinate u. For our purposes, the most helpful is the Gaussian gauge with the real distance l along the radial direction taken as a coordinate,

$$u \equiv l, \quad \alpha \equiv 0, \tag{6}$$

and the metric

$$ds^2 = e^{2\gamma(l)}\eta_{\mu\nu}dx^{\mu}dx^{\nu} - (dl^2 + e^{2\beta(l)}d\Omega^2). \tag{7}$$

Then, two independent components of the Einstein equations take the form (the prime now denotes d/dl)

$$\gamma'' + d_0\gamma'^2 + d_1\beta'\gamma' = -\frac{2\kappa^2}{D-2}V, \tag{8}$$

$$\beta'' + d_0\beta'\gamma' + d_1\beta'^2 = (d_1 - 1 - \kappa^2\phi^2)e^{-2\beta} - \frac{2\kappa^2}{D-2}V. \tag{9}$$

The Einstein equation

$$G^{l}_{l} = -\kappa^2 T^{l}_{l}$$

(where G^{B}_{A} is the Einstein tensor) is free of second-order derivatives:

$$\begin{aligned}&(d_0\gamma' + d_1\beta')^2 - d_0\gamma'^2 - d_1\beta'^2\\ &= \kappa^2(\phi'^2 - 2V) + d_1e^{-2\beta}(d_1 - 1 - \kappa^2\phi^2).\end{aligned} \tag{10}$$

The scalar field equations

$$\nabla^{A}\nabla_{A}\phi^{k} + \frac{\partial V}{\partial\phi^{k}} = 0$$

combine to yield an equation for $\phi(l)$:

$$\phi'' + (d_0\gamma' + d_1\beta')\phi' - d_1e^{-2\beta}\phi = \frac{dV}{d\phi}. \tag{11}$$

Due to the Bianchi identities, it is a consequence of Einstein equations (8)–(10). On the other hand, (10) is a first integral of Eqs. (8), (9), and (11).

In our analytical study, we do not specify any particular form of $V(\phi)$. However, we suppose that V has a maximum at $\phi = 0$ and a minimum at some $\phi = \eta > 0$, and, hence, $V'(0) = V'(\eta) = 0$. For convenience, we do not single out a cosmological constant, which may be identified with a constant component of the potential V or, in many cases, with its minimum value.

The parameter η (as the scalar field itself) has the dimension $[l^{-(D-2)/2}]$ and thus specifies a certain length scale $\eta^{-2/(D-2)}$ and energy scale $\eta^{2/(D-2)}$ (we use natural units such that $c = \hbar = 1$). In the conventional case $D = 4$, η has the dimension of energy and characterizes the SSB energy scale.

2.4. Regularity Conditions: A Regular Center

For the geometry to be regular, we must require finite values of all Riemann tensor components (2). In Gaussian gauge (6), the regularity conditions simply state that

$$\beta', \beta'', \gamma', \gamma'' \text{ are finite.} \tag{12}$$

For $d_1 > 0$, in addition to (12), a special regularity condition is needed at the center, which is a singular point of the spherical coordinates in $\mathbb{R}_u \times \mathbb{S}^{d_1}$. The center is a point where the radius $r \equiv e^{\beta}$ turns to zero. The regularity conditions there, also following from the finiteness of Riemann tensor components (2), are the same as in the usual static, spherically symmetric spacetime: in terms of an arbitrary u coordinate, they are given by

$$\begin{aligned}\gamma &= \gamma_c + O(r^2),\\ e^{\beta-\alpha}|\beta'| &= 1 + O(r^2) \text{ as } r \longrightarrow 0,\end{aligned} \tag{13}$$

where γ_c is a constant that can be set to zero by a proper choice of scales of the coordinates x^{μ}. The second condition in (13) follows, for $d_1 > 1$, from the finiteness of the Riemann tensor components R^{ab}_{cd} (see (2)). Its geometric meaning is the property of being locally Euclidean at $r = 0$, which implies that $dr^2 = dl^2$, i.e., the correct circumference-to-radius ratio for small circles. In the special case where $d_1 = 1$, with the quotient space $\mathbb{R}_u \times \mathbb{S}^{d_1}$ being two-dimensional, we obviously have $R^{ab}{}_{cd} \equiv 0$, but the second condition in (13) should still be imposed to avoid a conical singularity.

It is natural to put $l = 0$ at a regular center, then l is the distance from the center.

Regularity of the Ricci tensor components $R^{A}_{B} = R^{AC}{}_{BC}$ implies regularity of the stress–energy tensor T^{B}_{A}, whence it follows that

$$|V| < \infty, \quad e^{-\beta}|\phi| < \infty, \quad e^{-\alpha}|\phi'| < \infty \tag{14}$$

at any regular point and with any radial coordinate.

2.5. Boundary Conditions

Domain walls. For $d_1 = 0$, the metric in (1) or (7) describes a plane-symmetric five-dimensional spacetime, the coordinate l ranges over the entire real axis, and the broken symmetry is $\mathbb{Z}_2$, the mirror symmetry with respect to the plane $l = 0$. The topological defect is a domain wall separating two vacua corresponding to two values of a single real scalar field ϕ, e.g., $\phi = \pm\eta$. Accordingly, we assume that $\phi(l)$ is an odd function, whereas $\gamma(l)$ and $V(\phi)$ are even functions, and the conditions at $l = 0$ are

$$\gamma(0) = \gamma'(0) = \phi(0) = 0. \quad (15)$$

We thus have three initial conditions for the third-order set of equations (8) and (10) (Eq. (11) is their consequence), because the unknown function β is absent in this case.

Global strings and monopoles. For $d_1 > 0$, the regular center requirement leads to the following boundary conditions for Eqs. (8)–(10) at $l = 0$:

$$\phi(0) = \gamma(0) = \gamma'(0) = r(0) = 0, \quad r'(0) = 1. \quad (16)$$

We have five initial conditions for a fifth-order set of equations. However, $l = 0$, being a singular point of the spherical coordinate system (not to be confused with a spacetime curvature singularity), is also a singular point of our set of equations. As a result, the requirements of the theorem on the solution existence and uniqueness for our set of ordinary differential equations are violated. It turns out that the derivative $\phi'(0)$ remains undetermined by (16). If we set $\phi'(0) = 0$, we obtain a trivial (symmetric) solution with $\phi \equiv 0$ and a configuration without a topological defect. In the case where $V(0) = 0$, we arrive at the flat D-dimensional metric: we then have $\gamma \equiv 0$ and $r \equiv l$ in (7). If, however, $V(0) \neq 0$, the corresponding exact solutions to the Einstein equations for $d_0 > 1$, $d_1 > 1$ still need to be found. A direct inspection shows that it cannot be the de Sitter or AdS space: the constant curvature metrics are not solutions of the vacuum Einstein equations with a cosmological constant.

Nontrivial solutions exist if $\phi'(0) \neq 0$ and can correspond to SSB. We note that the very possibility of SSB appears as a result of violation of the solution's uniqueness at $r = 0$ provided that a maximum of the potential $V(\phi)$ at $\phi = 0$ corresponds to the center. The lacking boundary condition that may lead to a unique solution can now follow from the requirement of regularity at the other extreme of the range of l, whose nature is in turn determined by the shape of the potential.

In what follows, assuming a regular center, we try to find all possible conditions at the other extreme of the range $\mathbb{R}_l$ of the Gaussian radial coordinate, providing the existence of globally regular models with metric (7). In other words, we seek solutions with asymptotic forms such that the quantities in (2) are finite. Other regularity conditions, such as (14), then follow. In doing so, we do not restrict the possible shape of the potential $V(\phi)$ in advance. The cases under consideration are classified by the final values of $r = e^{\beta}$ (infinite, finite, or zero) and V (positive, negative, or zero). The scalar field ϕ is assumed to be finite everywhere.

Without loss of generality, we assume that $\phi'(0) > 0$ near $l = 0$, i.e., that ϕ increases, at least initially, as we recede from the center.

3. DOMAIN WALLS AND THE PROBLEM OF MATTER CONFINEMENT

Below, we mostly consider configurations with $d_1 \geq 2$ that correspond to a global monopole in the spherically symmetric space $\mathbb{R}_u \times \mathbb{S}^{d_1}$. Before that, we briefly discuss the problem of matter confinement on the brane and the complications involved in the five-dimensional case.

The metric coefficient $e^{2\gamma}$ in (1), sometimes called the warp factor, actually plays the role of a gravitational potential that determines an attractive or repulsive nature of gravity with respect to the brane. If it forms a potential well with a bottom on (or very near) the brane, there is the hope that matter, at least its low-energy modes, should be trapped.

It has been shown, in particular, that spin-1/2 fields are localized due to an increasing warp factor in (1 + 4)- and (1 + 5)-dimensional models [19, 20]. It has also repeatedly been claimed that, in (1 + 4) dimensions, a brane with an exponentially decreasing warp factor (as, e.g., in the RS2 model) can trap spin-0 and spin-2 fields. Our calculation for a scalar field shows that this is not the case.

A gravitational trapping mechanism suggested in [21] was characterized there as a universal one, suitable for all fields. It is based on nonexponential warp factors, which increase with distance from the brane and approach finite values at infinity. This mechanism was exemplified in [22] with a special choice of two so-called "smooth source functions" in the stress–energy tensor, describing a continuous distribution of certain phenomenological matter and vanishing outside the brane.

Our analysis uses more natural assumptions: a scalar field system admitting SSB, without any special choice of the symmetry-breaking potential, under the requirement of global regularity.

We briefly show, following [10, 11] (but in other coordinates), that this approach in (4 + 1) dimensions always leads to a decaying warp factor for any choice of $V(\phi)$ and that such a system cannot trap a test scalar field. We consider a domain wall in five dimensions, and, hence, $l \in \mathbb{R}$, we set $d_1 = 0$ in our equations, the

unknown $\beta(l)$ is absent, and Eqs. (8) and (11) for γ and the single scalar field ϕ are given by

$$\gamma'' + d_0\gamma'^2 = -\frac{2\kappa^2}{d_0-1}V, \tag{17}$$

$$\phi'' + d_0\gamma'\phi' - \frac{dV}{d\phi} = 0. \tag{18}$$

Their first integral in (10) reduces to

$$\gamma'^2 = -\frac{\kappa^2}{d_0(d_0-1)}(2V - \phi'^2). \tag{19}$$

The initial conditions at $l = 0$ corresponding to the $\mathbb{Z}_2$ symmetry (broken for the scalar field but preserved for the geometry) have form (15).

Eliminating V from (17) and (19) and integrating subject to (15), we obtain

$$(d_0-1)\gamma'(l) = -\kappa\int_0^l \phi'^2 dl, \tag{20}$$

and we conclude that $\gamma'(l)$ is negative at all $l > 0$ and describes gravitational repulsion from the brane; moreover, $e^{-\gamma}$ monotonically grows with growing l. The only possible regular solution corresponds to $|\gamma'(\infty)| < \infty$. Because $\gamma'(\infty) = 0$ in this case, it follows from Eq. (17) that $V(\infty) < 0$, corresponding to a negative cosmological constant $\Lambda = \kappa^2 V(\infty)$. Hence, the only possible regular asymptotic form is AdS, with

$$e^{\gamma} \approx ae^{-hl}, \quad a, h = \text{const}, \quad h = \sqrt{-\Lambda/6}. \tag{21}$$

The constant a depends on the particular shape of $V(\phi)$. At $l = \infty$, there is an AdS horizon ($e^{\gamma} = 0$), which, like a black hole horizon, attracts matter and prevents its trapping by the brane.

We show this for $d_0 = 4$ and a test scalar field χ with the Lagrangian

$$L_\chi = \frac{1}{2}\partial_A\chi^*\partial^A\chi - \frac{1}{2}m_0^2\chi^*\chi - \frac{1}{2}\lambda\phi^2\chi^*\chi, \tag{22}$$

where χ^* is the complex conjugate field and the last term describes a possible interaction between χ and the wall scalar field ϕ; λ is the coupling constant. The field $\chi(x^A)$ satisfies the linear homogeneous (modified Fock–Klein–Gordon) equation

$$\frac{1}{\sqrt{g}}\partial_A(\sqrt{g}g^{AB}\partial_B\chi) + (\lambda\phi^2 + m_0^2)\chi = 0. \tag{23}$$

Its coefficients depend on l only, and $\chi(x^A)$ may be sought in the form

$$\chi(x^A) = X(l)\exp(-ip_\mu x^\mu), \quad \mu = 0, 1, 2, 3, \tag{24}$$

where $p_\mu = (E, \mathbf{p})$ is a constant four-momentum. The function $X(l)$ determines the χ field distribution across the brane and satisfies the equation

$$X'' + 4\gamma'X' + [e^{-2\gamma}(E^2 - \mathbf{p}^2) - \lambda\phi^2 - m_0^2]X = 0. \tag{25}$$

The χ field is able to describe particles localized on the brane only if its stress–energy tensor $T_\mu^\nu[\chi]$ is finite in the whole five-space and decays sufficiently rapidly at large l. As an evident necessary condition of localization, the χ field energy per unit three-volume of the brane must be finite, i.e.,

$$\begin{aligned} E_{\text{tot}}[\chi] &= \int_{-\infty}^{\infty} T_t^t\sqrt{g}dl \\ &= \int_0^{\infty} e^{4\gamma}[e^{-2\gamma}(E^2 + \mathbf{p}^2)X^2 \\ &+ (m_0^2 + \lambda\phi^2)X^2 + X'^2]dl < \infty. \end{aligned} \tag{26}$$

Inequality (26) implies a finite norm of the χ field defined as

$$\|\chi\|^2 = \int_{-\infty}^{\infty}\sqrt{g}\chi^*\chi dl = \int_{-\infty}^{\infty} e^{4\gamma}X^2 dl. \tag{27}$$

At large l, because $e^{-2\gamma} \longrightarrow \infty$, the terms with λ and m_0 in Eq. (25) can be neglected and the equation determining the behavior of χ at large l can be written as

$$X'' - 4hX' + P^2e^{2hl}X = 0, \quad P^2 = \frac{E^2 - \mathbf{p}^2}{a^2h^2}. \tag{28}$$

It is solved in terms of Bessel functions, and the solution has the asymptotic form

$$X = Ce^{3hl/2}\sin(Pe^{hl} + \varphi_0), \quad z \longrightarrow \infty, \tag{29}$$

where C and φ_0 are integration constants. We see that quantity (29) is not only nonvanishing as $l \longrightarrow \infty$, but even oscillates with increasing amplitude. As a result, the stress–energy tensor components $T_\mu^\nu[\chi]$ are infinite at $l = \infty$. Moreover, integral (26) behaves as $\int e^{hl}dl$ and diverges. However, normalization integral (27) converges because the integrand behaves as e^{-hl}. This result is sometimes treated as a sufficient condition for localization, but, in our view, it is not true, because the very existence of the brane configuration is put in doubt if the test field stress–energy tensor is infinite somewhere.

Thus, a test scalar field with any mass tends to infinity as $l \longrightarrow \infty$ and develops an infinite stress–energy

tensor; even its interaction with the ϕ field that supports the brane does not improve the situation. We conclude that a single extra dimension is insufficient for providing gravitational attraction of matter to a regular isolated brane.

4. A SEARCH FOR REGULAR ASYMPTOTIC REGIMES

We now consider field equations (8)–(11) for global monopoles, assuming $d_1 \geq 2$. The string case $d_1 = 1$ is left aside, because it has some peculiarities that require a special study.

4.1. Solutions with the $r \longrightarrow \infty$ Asymptotic Regime

We define

$$\bar{V} = \frac{2\kappa^2 V}{D-2}, \quad \bar{V}_\infty = \bar{V}|_{r\to\infty}. \tag{30}$$

Evidently, $l \longrightarrow \infty$ as $r \longrightarrow \infty$, because, otherwise, we would have $\beta' \longrightarrow \infty$, violating the regularity conditions. The derivatives β' and γ' should tend to certain constant values to be denoted by β'_∞ and γ'_∞, respectively. Both β'' and γ'' vanish as $l \longrightarrow \infty$. Moreover, the second term in the right-hand side of Eq. (9) also vanishes. Therefore, in the leading order of magnitude, Eqs. (8) and (9) take the form

$$\begin{aligned}\gamma'_\infty(d_0\gamma'_\infty + d_1\beta'_\infty) &= -\bar{V}_\infty,\\ \beta'_\infty(d_0\gamma'_\infty + d_1\beta'_\infty) &= -\bar{V}_\infty.\end{aligned} \tag{31}$$

We consider the cases when $\bar{V}_\infty \neq 0$ and $\bar{V}_\infty = 0$ separately.

A1. $\bar{V}_\infty \neq 0$. Equations (31) immediately give

$$\beta'_\infty = \gamma'_\infty = \sqrt{-\bar{V}_\infty/(D-1)}, \quad \bar{V}_\infty < 0. \tag{32}$$

An evident necessary condition of the existence of regular configurations is $\bar{V}_\infty \leq 0$. We thus obtain

$$e^\beta \sim e^\gamma \sim e^{\beta'_\infty l},$$

and the metric takes the asymptotic form

$$ds^2 \approx C_1 e^{2\beta'_\infty l}\eta_{\mu\nu}dx^\mu dx^\nu - dl^2 - C_2 e^{2\beta'_\infty l}d\Omega^2, \tag{33}$$

with some positive constants C_1 and C_2. Equation (10) holds automatically if $\phi'(\infty) = 0$, as should be the case if we assume a finite asymptotic value of ϕ. Finally, in Eq. (11), all terms except $dV/d\phi$ manifestly vanish as $l \longrightarrow \infty$, and, hence, $dV/d\phi$ also vanishes, which should be the case if the field ϕ reaches an extremum of the potential V.

The finiteness condition for ϕ as $l \longrightarrow \infty$ separates a family of regular solutions among the continuum of integral curves leaving the regular center with different slopes $\phi'(0)$. As is confirmed by numerical experiments, if the potential has only one extremum (minimum) $\bar{V}_\infty < 0$, then there can be only one regular solution with $r \longrightarrow \infty$, $l \longrightarrow \infty$. However, there can be numerous regular solutions if the potential has several extremum points $\bar{V}_\infty < 0$.

In particular, if the initial maximum of the potential is located below the zero level, $V(0) \leq 0$, then there can be a continuum of regular integral curves starting from the regular center and returning to $\phi = 0$ at $l \longrightarrow \infty$. As can be verified numerically (see Section 4), there is a bunch of such curves parametrized by $\phi'(0) \in (0, \phi'_s)$, where $\phi'(0) = \phi'_s$ corresponds to a limiting regular curve (separatrix), also starting at $\phi(0) = 0$ but ending at the minimum $V(\eta)$.

The metric in (33) solves the Einstein equations with the stress–energy tensor $T_A^B = \delta_A^B V_\infty$ having the structure of a (negative) cosmological term. Moreover, according to (2), the Riemann tensor has the structure of a constant-curvature space at large l. In other words, such solutions have an anti-de Sitter (AdS_D) asymptotic form far from the center. But the metric in (33) is not a solution to our equations in the whole space even in the case where $\phi = \text{const}$. As already mentioned, for $d_0 > 1$ and $d_1 > 1$, constant-curvature metrics (dS_D and AdS_D) are not solutions of the vacuum Einstein equations with a cosmological constant.

A2. $\bar{V}_\infty = 0$.

Equations (31) are solved either by

$$\beta'_\infty = \gamma'_\infty = 0$$

or by

$$d_0\gamma'_\infty + d_1\beta'_\infty = 0.$$

However, when we substitute the second condition in Eq. (10), taking into account that $\phi' \longrightarrow 0$ at large l, we obtain

$$d_0\gamma'^2_\infty + d_1\beta'^2_\infty = 0$$

and return to

$$\beta'_\infty = \gamma'_\infty = 0.$$

Thus, both β' and γ' vanish at infinity, and we can try to seek them as expansions in inverse powers of l:

$$\beta' = \frac{\beta_1}{l} + \frac{\beta_2}{l^2} + \ldots, \quad \gamma' = \frac{\gamma_1}{l} + \frac{\gamma_2}{l^2} + \ldots. \tag{34}$$

Then, $O(l^{-2})$ is the leading order in the Einstein equations, and, to avoid contradiction,

$$r^{-2} = e^{-2\beta}$$

should be on the order of $O(l^{-2})$ or smaller. Moreover, because we assume that ϕ tends to a finite value $\phi_\infty > 0$, we have $\phi' = o(1/l)$ and scalar field equation (11) shows that

$$\frac{dV}{d\phi} = O(l^{-2})$$

or smaller; i.e., ϕ_∞ should be an extremum of $V(\phi)$. If $\phi(l)$ grows monotonically to $\phi_\infty > 0$, then ϕ_∞ is a minimum of V because, according to (11),

$$\frac{dV}{d\phi} < 0 \text{ as } \phi \longrightarrow \phi_\infty.$$

However, if $V(0) = 0$, one cannot exclude that ϕ returns to zero as $l \longrightarrow \infty$ (see item (c) below).

In the case where $\phi \longrightarrow \phi_\infty > 0$, because

$$V_\infty = dV/d\phi(\phi_\infty) = 0,$$

$V(\phi)$ is decomposed as

$$V(\phi) = \frac{1}{2}V_{\phi\phi}(\phi_\infty)(\phi - \phi_\infty)^2 + \ldots, \qquad (35)$$

where

$$V_{\phi\phi} = \frac{d^2V}{d\phi^2},$$

and, therefore,

$$V = o(l^{-2}).$$

As a result, Eqs. (8)–(10) lead to

$$\gamma_1(-1 + d_0\gamma_1 + d_1\beta_1) = 0, \qquad (36)$$

$$\beta_1(-1 + d_0\gamma_1 + d_1\beta_1) = \frac{l^2}{r^2}(d_1 - 1 - \kappa^2\phi_\infty^2), \qquad (37)$$

$$(d_0\gamma_1 + d_1\beta_1)^2 - d_0\gamma_1^2 - d_0\beta_1^2 = d_1\frac{l^2}{r^2}(d_1 - 1 - \kappa^2\phi_\infty^2). \qquad (38)$$

Now, it can be easily verified that we must necessarily set $\beta_1 = 1$. Indeed, for any $\beta_1 \neq 0$, we have

$$r = e^\beta \sim l^{\beta_1}.$$

Therefore, $\beta_1 < 1$ is excluded, because it leads to $r \ll l$, contrary to the above requirement. However, if we suppose that $\beta_1 > 1$, then $l^2/r^2 \longrightarrow 0$ as $l \longrightarrow \infty$, and Eq. (38) leads (if $\gamma_1 = 0$) to

$$\beta_1^2 = 0$$

or (if $\gamma_1 \neq 0$ and then $d_0\gamma_1 + d_1\beta_1 = 1$) to

$$d_0\gamma_1^2 + d_1\beta_1^2 = 1.$$

Both possibilities contradict the assumption that $\beta_1 > 1$.

Thus, $\beta_1 = 1$, and, hence,

$$r \approx kl, \quad k = \text{const} > 0,$$

at large l.

Equation (36) now leaves two possibilities,

$$\gamma_1 = 0$$

and

$$\gamma_1 = -\frac{d_1 - 1}{d_0},$$

and we consider them separately in items (a) and (b). Item (c) describes the case where expansions (34) do not work.

(a) If $\gamma_1 = 0$, then Eq. (37) yields

$$k^2 = 1 - \frac{\kappa^2\phi_\infty^2}{d_1 - 1}, \qquad (39)$$

and Eq. (10) in the same order is satisfied automatically. The metric takes the asymptotic form

$$ds^2 = e^{2\gamma_\infty}\eta_{\mu\nu}dx^\mu dx^\nu - dl^2 - k^2l^2d\Omega^2, \qquad (40)$$

where γ_∞ is a constant (we cannot turn it to zero by rescaling the coordinates x^μ, because such an operation has already been done for making $\gamma = 0$ at the center).

Thus, the whole metric has a flat asymptotic form, up to a solid angle deficit in the spherical part due to $k^2 \neq 1$. Such a deficit is a common feature of topological defects in the cases where they have (almost) flat asymptotic forms. Its appearance due to cosmic strings and global monopoles in spacetimes without extra dimensions is discussed in detail in [8]. For a global monopole in extra dimensions in the particular case where $d_0 = 4$ and $d_1 = 2$, it was treated by Benson and Cho [18]. We stress that the situation of a quasiflat asymptotic form with a solid angle deficit is not general. It occurs only for potentials with zero value at the minimum,

$$V(\phi_\infty) = 0,$$

and even in that case, not always, see item B below.

Namely, this geometry requires

$$|\phi_\infty| < \phi_{cr} := \frac{\sqrt{d_0 - 1}}{\kappa}, \tag{41}$$

i.e., ϕ_∞ should be smaller than the critical value ϕ_{cr} related to the D-dimensional Planck length. As ϕ_∞ approaches ϕ_{cr}, $k \longrightarrow 0$, the deficit absorbs the whole solid angle, and the above geometry disappears.

Scalar equation (11) shows how ϕ approaches ϕ_∞: in the leading order, we have

$$-\frac{d_1}{k^2 l^2} = V_{\phi\phi}(\phi_\infty)(\phi - \phi_\infty). \tag{42}$$

Assuming

$$V_{\phi\phi}(\phi_\infty) \neq 0,$$

we obtain

$$\phi - \phi_\infty \sim 1/l^2.$$

(b) If $\gamma_1 = -(d_1 - 1)/d_0$, then Eq. (37) leads to

$$\kappa^2 \phi_\infty^2 = d_1 - 1,$$

i.e.,

$$\phi_\infty = \phi_{cr},$$

while a substitution in (10) gives

$$(d_1 - 1)(d_0 + d_1 - 1) = 0,$$

contrary to our assumption that $d_1 - 1 > 1$. Therefore, this possibility does not lead to a regular asymptotic regime.

(c) If $V(0) = 0$, then a regular integral curve, starting at $l = 0$ and $\phi = 0$, can finish again with $\phi \longrightarrow 0$ as $l \longrightarrow \infty$. For large l and r, scalar field equation (11) for $|\phi| \ll 1$ reduces to

$$\phi'' + (d_0\gamma' + d_1\beta')\phi' - V_2\phi = 0, \tag{43}$$

where

$$V_2 = V_{\phi\phi}(0).$$

Because $\phi = 0$ is a maximum of $V(\phi)$ by assumption, we assume that $V_2 < 0$.

If we further assume that the function

$$s(l) = e^{d_0\gamma + d_1\beta}$$

satisfies the condition

$$s''/s \longrightarrow 0 \text{ as } l \longrightarrow \infty$$

(which is the case, e.g., for any power-behaved function), the solution of Eq. (43) is an oscillating function at large l,

$$\phi \approx \phi_0 e^{-(d_0\gamma + d_1\beta)/2} \cos[\sqrt{|V_2|}(l - l_0)], \quad l \longrightarrow \infty, \tag{44}$$

where ϕ_0 and l_0 are arbitrary constants. Substituting this in Eq. (8) and averaging $\cos^2 \longrightarrow 1/2$, we obtain

$$e^{d_0\gamma} \approx \frac{d_0\kappa^2|V_2|\phi_0^2}{2(D-2)} \int^l \frac{l\,dl}{r^{d_1}}, \quad l \longrightarrow \infty. \tag{45}$$

It is easy to verify that, for $d_1 > 2$, when the integral in (45) converges, the asymptotic form of the solution for $r = e^\beta$ and γ is $r \approx l$ and

$$\gamma = \gamma_\infty - \gamma_1/l^{d_1 - 2}, \quad \gamma_1, \gamma_\infty = \text{const},$$

i.e., we have a flat asymptotic regime.

In the special case where $d_1 = 2$, the integral diverges logarithmically and the solution may be approximated as (again)

$$r \approx l$$

and

$$e^\gamma \approx \text{const} \cdot \ln l.$$

This "logarithmic" asymptotic from resembles the behavior of cylindrically symmetric solutions in standard general relativity.

4.2. Solutions with the $r \longrightarrow r_* > 0$ Asymptotic Regime

Evidently, a regular solution cannot terminate at finite r and $l < \infty$. Therefore, we seek a regular asymptotic regime as $l \longrightarrow \infty$, where r and ϕ tend to finite limits, r_* and ϕ_*, and, hence, the quantities β', β'', ϕ', and ϕ'' vanish.

Moreover, in a regular solution, γ' should tend to a finite limit as $l \longrightarrow \infty$, and, hence, $\gamma'' \longrightarrow 0$. As a result, Eqs. (8) and (9) at large l lead to

$$d_0\gamma'^2 = -\bar{V}_* = \frac{1}{r_*^2}(\kappa^2\phi_*^2 - d_1 + 1), \tag{46}$$

where $\bar{V}_* = \bar{V}(\phi_*)$. We see that $\bar{V}_* \leq 0$ and, in addition, the scalar field should be critical or larger, $\phi_* \geq \phi_{cr}$. According to (46), at large l,

$$\pm\gamma' \approx h := \sqrt{-\bar{V}_*/d_0} \geq 0, \tag{47}$$

and Eq. (10), as in the previous cases, simply verifies that the solution is correct in the leading order. The sca-

lar field equation gives a finite asymptotic value of $V_\phi \equiv dV/d\phi$:

$$V_\phi(\phi_*) = -d_1\phi_* r_*^{-2}. \quad (48)$$

This value is negative if $\phi_* > 0$.

We obtain different asymptotic regimes for negative, positive, and zero values of γ'.

B1. $e^\gamma \sim e^{-hl}$, $h > 0$. The metric has the asymptotic form

$$ds^2 = C^2 e^{-2hl}\eta_{\mu\nu}dx^\mu dx^\nu - dl^2 - r_*^2 d\Omega^2. \quad (49)$$

The extra-dimensional part of the metric again describes an infinitely long cylindrical tube, but now the vanishing function $g_{tt} = e^{2\gamma}$ resembles a horizon. The substitution $e^{-hl} = \rho$ (converting $l = \infty$ to a finite coordinate value, $\rho = 0$) brings metric (49) to the form

$$ds^2 = C^2\rho^2\eta_{\mu\nu}dx^\mu dx^\nu - \frac{d\rho^2}{h^2\rho^2} - r_*^2 d\Omega^2. \quad (50)$$

Thus, $\rho = 0$ is a second-order Killing horizon in the two-dimensional subspace parametrized by t and ρ; it is of the same nature as, e.g., the extreme Reissner–Nordström black hole horizon, or the AdS horizon in the second Randall–Sundrum braneworld model. A peculiarity of the present horizon is that the spatial part of the metric, which at large l takes the form $\rho^2(d\mathbf{x})^2$, is degenerate at $\rho = 0$. The volume of the d_0-dimensional spacetime vanishes as $l \longrightarrow \infty$, and it remains degenerate even if we pass to Kruskal-like coordinates in the (t, ρ) subspace. However, the D-dimensional curvature is finite there, indicating that the transition to negative values of ρ (where the old coordinate l no longer works) is meaningful.[1]

One more observation can be made. According to (46), the potential V is necessarily negative at large l. On the other hand, Eq. (8) may be rewritten in integral form:

$$e^{d_0\gamma + d_1\beta}\gamma' = -\int_0^l e^{d_0\gamma + d_1\beta}\bar{V}dl. \quad (51)$$

The lower limit of the integral corresponds to a regular center, where the left-hand side of (51) vanishes. As $l \longrightarrow \infty$, it also vanishes due to $\gamma \longrightarrow -\infty$. Thus, the integral in the right-hand side, taken from zero to infinity, is zero. This means that the potential $V(\phi)$ has alternate sign and is positive in a certain part of the range $(0, \phi_*)$.

Thus, purely scalar solutions of the monopole type may contain second-order horizons. The degenerate nature of the spatial metric at the horizon does not lead to a curvature singularity, and the solutions may be continued in a Kruskal-like manner. Nevertheless, we do not consider these solutions as ones describing viable monopole configurations, because the zero volume of the corresponding spatial section makes the density of any additional (test) matter infinite. It is then impossible to neglect its backreaction, which evidently destroys such a configuration.

B2. $e^\gamma \sim e^{hl}$, $h > 0$. The metric has the asymptotic form

$$ds^2 = C^2 e^{2hl}\eta_{\mu\nu}dx^\mu dx^\nu - dl^2 - r_*^2 d\Omega^2, \quad C = \text{const} > 0. \quad (52)$$

Thus, in the spherically symmetric extradimensional part of the metric, we have an infinitely long d_1-dimensional cylindrical "tube" with an infinitely growing gravitational potential $g_{tt} = e^{2\gamma}$.

With this cylindrical asymptotic form, according to (47) and (48), the potential V tends to a negative value and has a negative slope. Moreover, the integral in Eq. (51) is negative and diverges at large l due to growing e^γ.

Regular solutions with $\gamma'(\infty) > 0$ naturally arise if the potential $V(\phi)$ is negative everywhere. We note, however, that, when $V(0)$ is above zero, by (51), the function $\gamma(l)$ decreases near the center ($l = 0$) due to $V > 0$ and grows at large l. It therefore has a minimum at some $l > 0$.

B3. $\bar{V}_* = 0$. This case contains one more asymptotic regime where the extra space ends with a regular tube.

Indeed, we can once again use expansions (34), but now with ϕ_* instead of ϕ_∞ and $\beta_1 = 0$ in accordance with $r \longrightarrow r_*$. Equation (9) (order $O(1)$) shows that

$$\kappa^2\phi_*^2 = d_1 - 1,$$

[1] One might wonder why we here do not obtain simple (first-order) horizons, like those in the Schwarzschild and de Sitter metrics, while such horizons generically appeared in the special case $d_0 = 1$, which corresponds to spherically symmetric global monopoles in general relativity, considered in detail in [14, 16].

The reason is that for $d_0 = 1$, $\delta^{\mu\nu}_{\rho\sigma}$ in (3) is zero, and the corresponding component of the Riemann tensor is also zero regardless of the values of γ'. In terms of the Gaussian coordinate l, a simple horizon occurs at some finite $l = l_h$, near which $g_{tt} = e^{2\gamma} \sim (l - l_h)^2$, such that $\gamma' \longrightarrow \infty$. When $d_0 = 1$, this does not lead to a singularity, because only the combinations $\gamma'' + \gamma'^2$ and $\beta'\gamma'$ are then required to be (and actually are) finite. In the case where $d_0 > 1$, instead of a horizon, we would have a curvature singularity at finite l, a situation excluded from the present study.

We thus have a general result for the metric in (1): for $d_0 > 1$, horizons can only be of order 2 or higher.

i.e.,

$$\phi_* = \phi_{cr}.$$

Equation (11) (order $O(1)$) gives a finite value of the derivative

$$dV/d\phi(\phi_*) = -d_1\phi_*/r_*^2.$$

Further, Eq. (8) (order $O(l^{-2})$) yields

$$\gamma_1(d_0\gamma_1 - 1)/l^2 = -\bar{V},$$

showing that

$$V = O(l^{-2})$$

(or even smaller). Because

$$V = (dV/d\phi(\phi_*))(\phi - \phi_*) + o(\phi - \phi_*),$$

we have to conclude that

$$\phi - \phi_* = O(l^{-2})$$

or smaller.

Now, assuming

$$V(\phi) = V_2/l^2 + \ldots,$$

we can find V_2 directly as the leading term in

$$(dV/d\phi(\phi_*))(\phi - \phi_*)$$

and, independently, from Eq. (9) (order $O(l^{-2})$), obtaining the two expressions

$$V_2 = -d_1\frac{\phi_*\phi_2}{r_*^2}$$

and

$$V_2 = -(D-2)\frac{\phi_*\phi_2}{r_*^2},$$

whence it follows that $d_1 = D - 2$, or $d_0 = 1$. Such a "critical" asymptotic regime ($\phi \longrightarrow \phi_{cr}$, $g_{tt} \longrightarrow 0$, and $r \longrightarrow$ const) was indeed found for $d_0 = 1$ in our papers [14, 16] describing $(d_1 + 2)$-dimensional spherically symmetric global monopoles, but, as we see, it does not exist in the case under consideration, $d_0 > 1$.

The only remaining possibility is that

$$\phi - \phi_* = o(l^{-2})$$

and

$$\gamma \longrightarrow \gamma_* = \text{const},$$

i.e., a solution tending, at large l, to the following simple "flux-tube" solution, valid for any d_0 and d_1:

$$r = \text{const}, \quad \gamma = \text{const}, \quad \phi = \phi_{cr}, \quad V = 0, \quad dV/d\phi = -d_1\phi_{cr}/r^2. \tag{53}$$

Such a solution can exist if the potential $V(\phi)$ has the properties

$$V(\phi_{cr}) = 0$$

and

$$dV/d\phi(\phi_{cr}) < 0,$$

and the last equality in (53) then relates the constant radius r to $dV/d\phi(\phi_{cr})$.

4.3. Solutions with the $r \longrightarrow 0$ Asymptotic Regime

The limit $r \longrightarrow 0$ means a center, and for it to be regular, conditions (12) must hold, and, hence, for our system, initial conditions (16) with $l = 0$ should be replaced, e.g., with $l = l_0 > 0$.

We now recall that conditions (16) determine the solution to the field equations for a given potential $V(\phi)$ up to the value of ϕ'. In particular, if there is one more center at $l = l_0$, then, starting from it and choosing

$$\phi'(l_0) = -\phi'(0),$$

we obtain the same solution in terms of $l_0 - l$ instead of l. We thus obtain a solution with two regular centers that is symmetric with respect to the middle point $l = l_0/2$, to be called the equator. To be smooth there, it must satisfy the conditions

$$\beta' = \gamma' = \phi' = 0 \text{ at } l = l_0/2, \tag{54}$$

which implicitly restrict the shape of the potential. Given a potential $V(\phi)$, conditions (54) create, in general, three relations among l_0, $\phi'(0)$, and the free parameters of $V(\phi)$(if any). Eliminating l_0 and $\phi'(0)$, we must obtain a single "fine tuning" condition for the parameters of the potential.

A necessary condition for the existence of such a solution is that $V(\phi)$ has a variable sign. This follows from Eq. (51) by integration over the segment $(0, l_0/2)$: the integral vanishes because $\gamma' = 0$ at both ends.

Moreover, as follows from Eqs. (9) and (10) with (54),

$$r_e^{-2}(d_1 - 1 - \kappa^2\phi_e^2) = \frac{D-2}{d_1}\bar{V}_e = \beta_e'' + \bar{V}_e, \tag{55}$$

Classification of global monopole solutions for arbitrary $V(\varphi)$ by asymptotic types. Attraction or repulsion is understood with respect to the center

Notation	r	$V(\varphi)$	φ	γ	Asymptotic type
A1	∞	$V(\eta) < 0$	$\eta < \varphi_{cr}$	∞	AdS, attraction
A2(a)	∞	0	$\eta < \varphi_{cr}$	const	Flat, solid angle deficit
A2(c), $d_1 > 2$	∞	0	0	const	Flat
A2(c), $d_1 = 2$	∞	0	0	∞	"Logarithmic", attraction
B1	r_*	$V_* < 0$	$\varphi_* > \varphi_{cr}$	$-\infty$	Double horizon, repulsion
B2	r_*	$V_* < 0$	$\varphi_* > \varphi_{cr}$	∞	Attracting tube
B3	r_*	0	$\varphi_* = \varphi_{cr}$	const	Trivial tube
C	0	$V(0)$	0	const	Second center

leading to

$$d_1\beta_e'' = (d_0 - 1)\bar{V}_e$$

(where the index "e" refers to values at the equator). If $r = e^{\beta}$ is assumed to grow monotonically from zero to its maximum value at the equator, we have $\beta_e'' < 0$. Hence, $\bar{V}_e < 0$, and (55) implies that $\phi_e > \phi_{cr}$; i.e., the scalar field at the equator must exceed its critical value.

The existence of asymmetric solutions with two regular centers, corresponding to

$$\phi'(l_0) \neq \phi'(0),$$

is also possible. In this case, there would be no equator in general, because β and ϕ would have maxima at different l; moreover, we would have in general,

$$\gamma(l_0) \neq \gamma(0) = 0,$$

and $\gamma(l)$ could even have no extremum. However, because $\gamma' = 0$ at both centers, the integral in (51) taken from 0 to l_0 should vanish and hence, again, V would have alternating sign.

The whole configuration with two regular centers has the topology $\mathbb{M}^{d_0} \times \mathbb{S}^{d_1+1}$, with closed extra dimensions in the spirit of Kaluza–Klein models. The main difference from them is that all variables now essentially depend on the extra coordinate l.

The main properties of all regular asymptotic regimes found, which lead to a classification of possible global monopole configurations in extra dimensions, are summarized in the table. The word "attraction" corresponds to an increasing warp factor far from the brane.

5. SCALAR FIELD TRAPPING BY GLOBAL MONOPOLES

We consider a test scalar field with Lagrangian (22) in the background of global monopole configurations described in Section 4. After variable separation (24), the field equation for a **p**-mode of the scalar field χ becomes

$$X'' + (d_0\gamma' + d_1\beta')X' + (e^{-2\gamma}p^2 - \mu^2)X = 0, \quad (56)$$

where

$$p^2 = p_\mu p^\mu = E^2 - \mathbf{p}^2$$

is the d_0-momentum squared and

$$\mu^2 = m_0^2 + \lambda\phi^2$$

is the effective mass squared. The trapping criterion consists, as before, in the requirements that the χ field stress–energy tensor must vanish far from the brane and the total χ field energy per unit volume of the brane must be finite, i.e.,

$$E_{\mathrm{tot}}[\chi] = \int \sqrt{g}\, d^{d_1+1}x \times [e^{-2\gamma}(E^2 + \mathbf{p}^2)X^2 + \mu^2X^2 + X'^2]dl < \infty. \quad (57)$$

The first requirement means that each term in the square brackets in (57) must vanish at large l.

We now check whether these requirements can be met at different kinds of asymptotic regimes listed in the table.

A1. Attracting AdS asymptotic regime $\beta \sim \gamma \sim kl$, $k > 0$. At large l, Eq. (56) reduces to the equation with constant coefficients

$$X'' + (D-1)X - \mu^2 X = 0,$$

and its solution vanishing as $l \longrightarrow \infty$ is

$$X \sim e^{-al},$$
$$a = \frac{1}{2}[(D-1)k + \sqrt{(D-1)^2k^2 + 4\mu^2}]. \quad (58)$$

It can be directly verified that the trapping requirements are satisfied for all momenta $\mathbf{p}$ and all $\mu^2 \geq 0$.

A2(a). A quasiflat asymptotic regime with a solid angle deficit. At large l, Eq. (25) reduces to

$$X'' + d_1 X/l + P^2 X = 0,$$

where

$$P^2 = p^2 e^{-2\gamma_\infty} - \mu^2$$

and γ_∞ is the limiting value of γ at $l = \infty$. In terms of

$$Y = l^{d_1/2} X,$$

this equation is (at large l) rewritten as

$$Y'' + P^2 Y = 0,$$

while trapping condition (57) implies that

$$\int l^{d_1} X^2(l)\, dl < \infty.$$

Therefore, only an exponentially falling $Y(l)$ is suitable. In other words, the trapping condition is $P^2 < 0$, or

$$p^2 < m_{\rm cr}^2 := \mu^2 e^{2\gamma_\infty}, \quad (59)$$

where, now,

$$\mu^2 = m_0^2 + \lambda^2 \eta^2.$$

We note that

$$p^2 = E^2 - \mathbf{p}^2$$

is nothing else but the observable mass of a free χ-particle if the observer watches its motion in the Minkowski section $l = 0$ of our manifold, i.e., on the brane. Hence, condition (59) means that the brane traps all scalar particles of masses smaller than the critical value $m_{\rm cr}$ depending on the model parameters.

A2(c) ($d_1 > 2$). This case differs from the previous one only by the asymptotic value of ϕ, which is now zero, and, hence, $\mu = m_0$.

A2(c) ($d_1 = 2$). A "logarithmic" asymptotic regime, $e^\gamma \sim \ln l$. Because $e^{-2\gamma} \sim 1/(\ln l)^2 \longrightarrow 0$, the term with p^2 drops out of Eq. (56), which then leads to the decreasing solution

$$X \sim l^{-1} e^{-\mu l},$$

and a χ-particle is trapped provided $\mu = m_0 > 0$.

B1. A horizon. As was remarked previously, we do not regard this configuration viable and omit it from our discussion.

B2. An attracting tube, $r \longrightarrow r_*$ and $\gamma \approx hl$, $h > 0$ as $l \longrightarrow \infty$. Equation (56) takes the form

$$X'' + d_0 h X' - \mu^2 X = 0$$

and has the decreasing solutions

$$X \sim e^{-al}, \quad a = \frac{1}{2}(d_0 h + \sqrt{d_0^2 h^2 + 4\mu^2}). \quad (60)$$

As in item A1, it is easy to verify that the trapping conditions hold provided $\mu^2 > 0$.

B3. A trivial tube, both β and γ tend to constants as $l \longrightarrow \infty$. In Eq. (56), the term with X' drops out at large l, and an exponentially decreasing solution exists under condition (59), where

$$\mu^2 = m_0^2 + \lambda^2 \phi_{\rm cr}^2.$$

C. These configurations have no large l asymptotic regimes and are not interpreted in terms of branes.

A conclusion is that scalar particles of any mass and momentum are trapped by global monopoles with A1 and B2 asymptotic regimes with exponentially growing warp factors and A2(c) with a logarithmic asymptotic regime; they are trapped under restrictions (59) on the particle's observable mass by monopoles with A2 and B3 asymptotic regimes whose warp factors tend to constant limits far from the brane.

6. NUMERICAL RESULTS: MEXICAN HAT POTENTIAL

In this section, we present the results of our numerical calculations, which confirm the classification of regular solutions given above. We have used the Mexican hat potential in the form (Fig. 1)

$$V = \frac{\lambda\eta^4}{4}\left[\varepsilon + \left(1 - \frac{\phi^2}{\eta^2}\right)^2\right]. \quad (61)$$

It has two extremum points in the range $\phi \geq 0$: a maximum at $\phi = 0$ and a minimum at $\phi = \eta$. The SSB energy

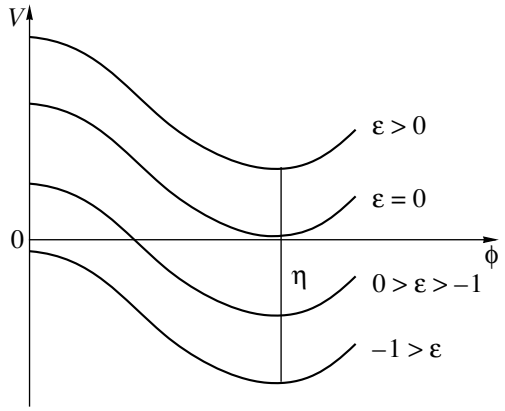


Fig. 1. Mexican hat potential.

scale is characterized by $\eta^{2/(D-2)}$, while $\sqrt{\lambda}\eta$ determines, as usual, the length scale. The nonconventional parameter ε introduced in (61) moves the potential up and down, which is equivalent to adding a cosmological constant to the usual Mexican hat potential.

Given potential (61), the nature of the solutions essentially depends on its two dimensionless parameters: ε, fixing the extremal values of the potential with respect to zero, and $\kappa^2\eta^2$, characterizing the gravitational field strength: as we remember from Section 4, the asymptotic regime $r \longrightarrow \infty$ only exists when $\phi_\infty < \phi_{cr}$, which is the same as

$$\kappa^2\eta^2 < d_1 - 1.$$

If ε > 0, potential (61) is always positive, and, in accordance with item A1, regular solutions are absent.

In the conventional case where ε = 0, in the range

$$0 < \kappa^2\eta^2 < d_1 - 1,$$

there are asymptotically flat regular solutions with a solid angle deficit (class A2).

The most complex case 0 > ε > –1 contains a variety of possibilities. Regular solutions with the asymptotic behavior $r \longrightarrow \infty$ as $l \longrightarrow \infty$ having $\gamma'_\infty > 0$ (case A1) exist in some range $0 < \eta < \eta_s$, where the separating value η_s depends on d_0, d_1, and ε. As an example, such a regular solution with $\kappa^2\eta^2 = 5$, ε = –0.75, $d_0 = 4$, and $d_1 = 3$ is presented in Fig. 2.

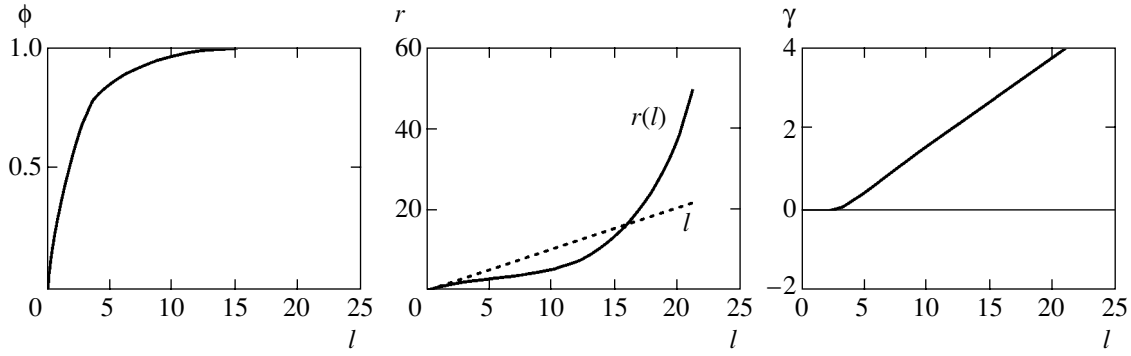


Fig. 2. A regular solution with an AdS asymptotic regime (type A1) for potential (61) with $\kappa\eta^2 = 5$, ε = –0.75, $d_0 = 4$, $d_1 = 3$.

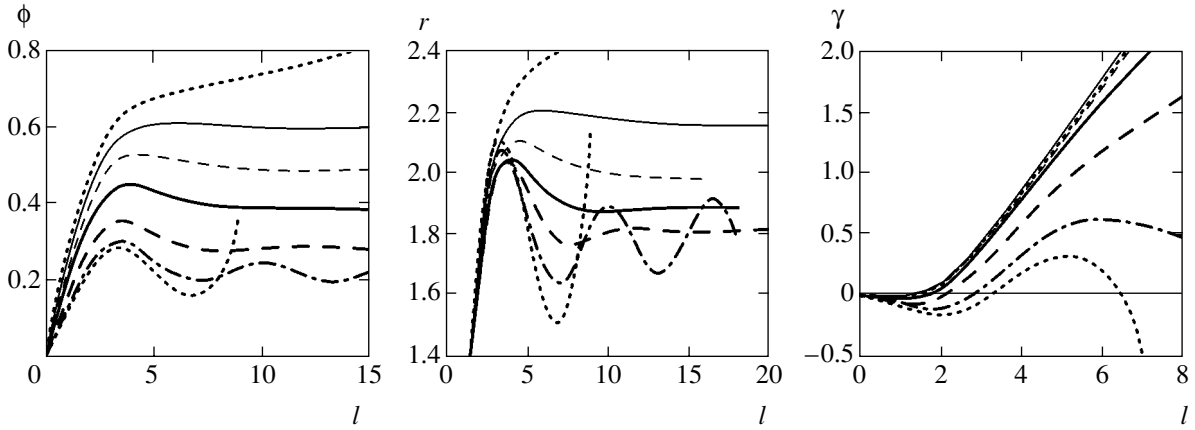


Fig. 3. Regular (except for dotted curves) solutions with the B2 asymptotic regime (attracting tube), such that $r \longrightarrow r_* < \infty$ and $\gamma'_\infty > 0$.

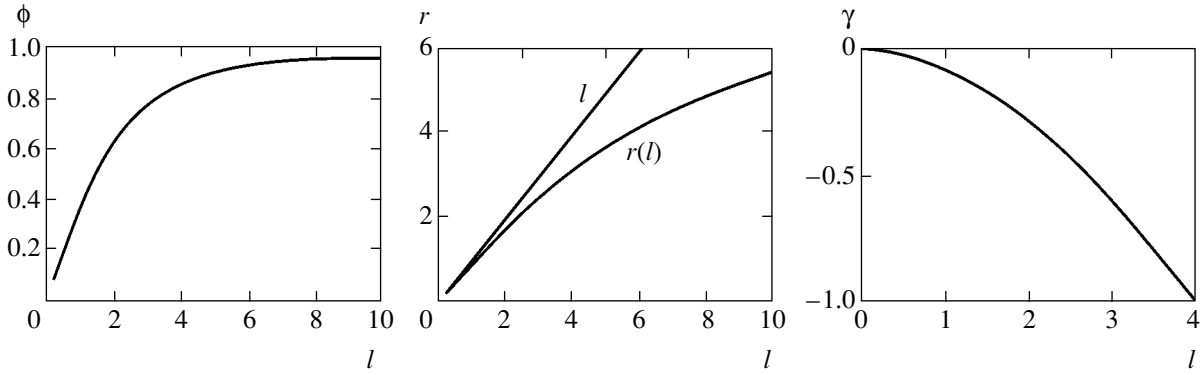


Fig. 4. A regular solution with the asymptotic regime $r \longrightarrow r_* < \infty$ and $\gamma'_\infty < 0$ (case B1, horizon).

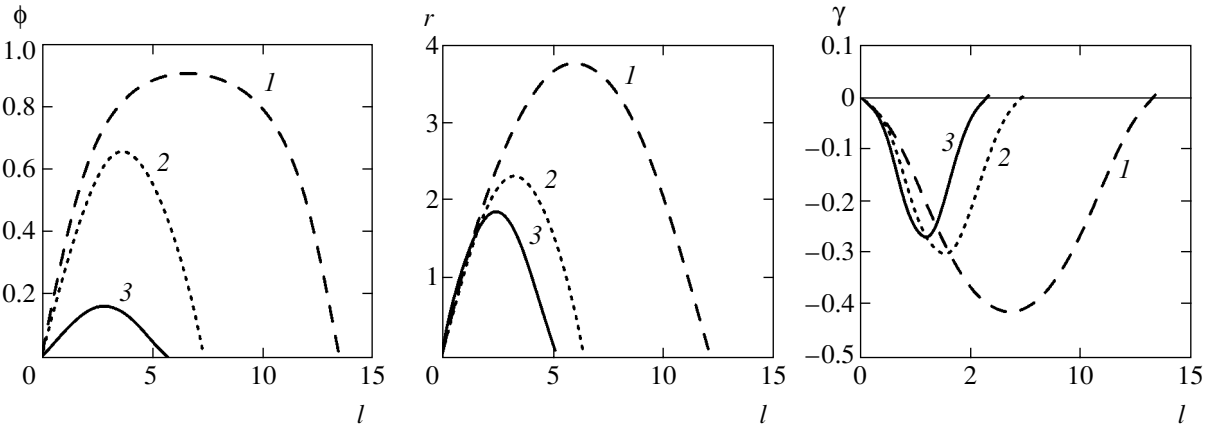


Fig. 5. Type-C solutions with two regular centers ($r \longrightarrow 0$, $\phi \longrightarrow 0$, $\gamma' \longrightarrow 0$ as $l \longrightarrow l_0$).

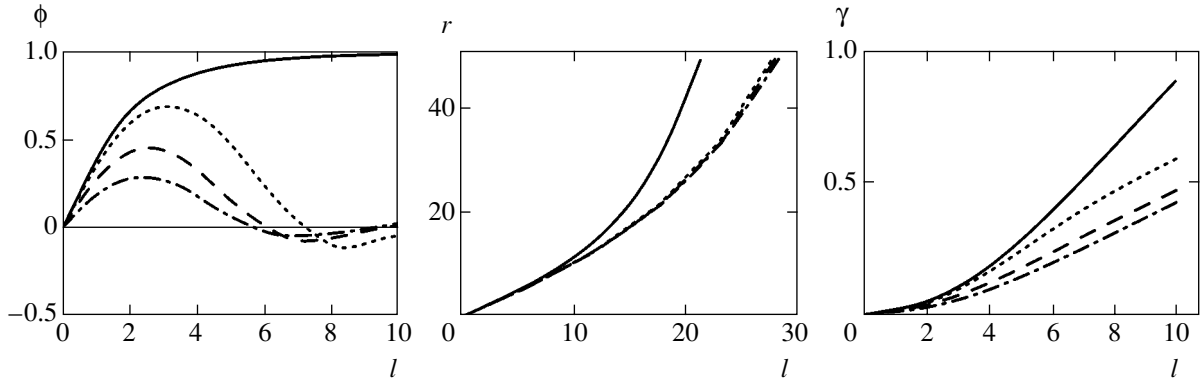


Fig. 6. Regular solutions starting and terminating at $\phi(0) = \phi(\infty) = 0$. The limiting solid curve with $\phi'(0) = \phi'_c = 0.4401425$ (separatrix) terminates at $\phi(\infty) = \eta$.

Depending on the parameters of the potential, there are regular solutions with the asymptotic regime $r \longrightarrow r_* < \infty$ and $\gamma'_\infty > 0$ (case B2) in a certain range $\eta_{s1} < \eta < \eta_{s2}$ (see Fig. 3). Here, $\varepsilon = -0.9$, $d_0 = 4$, $d_1 = 3$. The curves are given for $\kappa^2\eta^2 = 10$, 12, 15, 20, 30, 40, and 45 (from top down). The dotted curves ($\kappa^2\eta^2 = 10$ and $\kappa^2\eta^2 = 45$) correspond to singular configurations. It follows that, for $\varepsilon = -0.9$, $d_0 = 4$, $d_1 = 3$, the lower bound of this parameter that leads to regular models is somewhere between 10 and 12, while the upper bound is between 30 and 45.

An example of a regular solution with the asymptotic regime $r \longrightarrow r_* < \infty$ and $\gamma'_\infty < 0$ (class B1), corre-

sponding to a second-order Killing horizon, is shown in Fig. 4.

The value $\kappa\eta^2 = 17.37$ is fine-tuned to the parameters $\varepsilon = -0.75$, $d_0 = 4$, $d_1 = 2$ of this particular solution.

Other examples of fine-tuned regular solutions, namely, type C with two regular centers ($r \longrightarrow 0, \phi \longrightarrow 0, \gamma' \longrightarrow 0$ at $l \longrightarrow l_0$), are presented in Fig. 5.

For all three curves, $d_0 = 4$ and $d_1 = 2$. The curves (*1*, *2*, and *3*) correspond to $\varepsilon = -0.15$, -0.5, and -0.9626, respectively. The fine-tuned values of $\kappa^2\eta^2$ are approximately 2.637, 6.17, and 100.

In the case $\varepsilon \le -1$, the maximum $V(0) \le 0$ is at or below the zero level, and there is a possibility for the integral curves to start and finish at the same value $\phi(0) = \phi(\infty) = 0$. We then observe a whole family of such regular curves in the range $0 < \phi'(0) < \phi'_c$ (see Fig. 6).

For the particular example presented ($\varepsilon = -1.5$, $\kappa\eta^2 = 1$, $d_0 = 4$, $d_1 = 3$), the values of $\phi'(0)$ for the dotted curves ending with $\phi = 0$ are 0.2, 0.3, and 0.4 (from bottom up). The limiting solid curve with $\phi'(0) = \phi'_c = 0.4401425$ (separatrix) is a regular solution ending at the minimum of the potential: $\phi \longrightarrow \eta$ as $l \longrightarrow \infty$.

The Mexican hat potential (61), with its only two extrema at $\phi = 0$ and $\phi = \eta$, cannot demonstrate the whole variety of solutions that appear with more sophisticated potentials having additional maxima and/or minima. Thus, for instance, class-A solutions may have a large-r asymptotic regime at any such extremum.

7. CONCLUSIONS

We have obtained as many as seven classes of regular solutions of the field equations describing a Minkowski thick brane with a global monopole in extra dimensions (see table).

Some of these classes, namely, A1 with an AdS asymptotic behavior and B2 ending with an attracting tube, have the exponentially growing warp factor $e^{2\gamma}$ at large l and are shown to trap linear test scalar fields modes of any mass and momentum.

Others, A2(a) and A2(c) for $d_1 > 2$, ending with a flat metric at large l, have a warp factor tending to a constant whose value is determined by the shape of the potential $V(\phi)$. They are also shown to trap a test scalar field, but the observable mass of the field is restricted from above by a value depending on the particular model of the global monopole.

Lastly, for $d_1 = 2$, i.e., a three-dimensional global monopole in the extra dimensions, class A2(c) solutions have a logarithmically growing warp factor. All test scalar field modes are trapped by this configuration, but the slow growth of $\gamma(l)$ probably means that the test field is strongly smeared over the extra dimensions.

All such configurations, in sharp contrast to RS2-like domain walls in five dimensions, are able to trap scalar matter. It is certainly necessary to check whether nonzero-spin fields are trapped as well and Newton's law of gravity holds on the brane in conformity with the experiment. We hope to consider these subjects in our future publications.

In addition to the trapping problem, a shortcoming of RS2-type Minkowski branes is that they are necessarily fine-tuned. Many of the global monopole solutions, at least those existing in the weak gravity regime (class A), are free of this shortcoming and are thus better for thick brane model building.

Some results and conclusions of this paper were previously given in [12, 13]. The main difference of our approach from theirs is their boundary condition, which is $\phi = \eta$ in our notation. This excludes the cases where the solution ends at a maximum or slope of the potential, such as symmetric solutions with two regular centers. Another difference is that they consider solutions with an exponentially decreasing warp factor as those leading to matter confinement on the brane. In our view, such solutions with second-order horizons do not represent viable models of a braneworld. We conclude that the present paper gives the most complete classification of all regular solutions for global monopoles in extra dimensions, which, even without gauge fields, seem to be promising as braneworld models.

ACKNOWLEDGMENTS

This work was supported, in part, by the Russian Foundation for Basic Research, project no. 05-02-17484. One of us (K.B.) gratefully acknowledges partial support under ISTC Project 1655 and DFG grant 436 RUS 113/807/0-1(R).

REFERENCES

1. K. Akama, Lect. Notes Phys. **176**, 267 (1982); V. A. Rubakov and M. E. Shaposhnikov, Phys. Lett. B **125**, 136 (1983); M. Visser, Phys. Lett. B **159**, 22 (1985); I. Antoniadis, Phys. Lett. B **246**, 377 (1985); M. Pavšič, Phys. Lett. A **116**, 1 (1986); gr-qc/0101075; Nuovo Cimento A **95**, 297 (1986); E. J. Squires, Phys. Lett. B **167**, 286 (1986); G. W. Gibbons and D. L. Wiltshire, Nucl. Phys. B **287**, 717 (1987).
2. P. Horava and E. Witten, Nucl. Phys. B **460**, 506 (1996); Nucl. Phys. B **475**, 94 (1996); E. Witten, Nucl. Phys. B **471**, 135 (1996).
3. V. A. Rubakov, Usp. Fiz. Nauk **171**, 913 (2001) [Phys. Usp. **44**, 871 (2001)]; F. Quevedo, Class. Quantum Grav. **19**, 5721 (2002); S. Nojiri, S. D. Odintsov, and S. Ogushi, Int. J. Mod. Phys. A **17**, 4809 (2002); P. Brax and C. van de Bruck, hep-th/0303095; R. Maartens, Living Rev. Rel. **7**, 7 (2004); gr-qc/0312059; P. Kanti, Int. J. Mod. Phys. A **19**, 4899 (2004); hep-ph/0402168.
4. L. Rundall and R. Sundrum, Phys. Rev. Lett. **83**, 4690 (1999).

5. D. A. Kirzhnits, Pis'ma Zh. Éksp. Teor. Fiz. **15**, 745 (1972) [JETP Lett. **15**, 529 (1972)].
6. Ya. B. Zel'dovich, I. Yu. Kobzarev, and L. B. Okun', Zh. Éksp. Teor. Fiz. **67**, 3 (1974) [Sov. Phys. JETP **40**, 1 (1975)].
7. A. H. Guth, Phys. Rev. D **23**, 347 (1981); A. D. Linde, Phys. Lett. B **114**, 431 (1982); V. A. Berezin, V. A. Kuzmin, and I. I. Tkachev, Phys. Lett. B **120**, 91 (1983); **124**, 479 (1983).
8. A. Vilenkin and E. P. S. Shellard, *Cosmic Strings and Other Topological Defects* (Cambridge Univ. Press, Cambridge, 1994).
9. C. Csaki, J. Erlich, T. J. Hollowood, and Y. Shirman, Nucl. Phys. B **581**, 309 (2000); D. Bazeia, F. A. Brito, and J. R. Nascimento, Phys. Rev. D **68**, 085007 (2003); S. Kobayashi, K. Koyama, and J. Soda, Phys. Rev. D **65**, 064014 (2002); A. Melfo, N. Pantoja, and A. Skirzewski, Phys. Rev. D **67**, 105003 (2003).
10. K. A. Bronnikov and B. E. Meierovich, Gravit. Cosmol. **9**, 313 (2003); gr-qc/0402030.
11. S. T. Abdyrakhmanov, K. A. Bronnikov, and B. E. Meierovich, Gravit. Cosmol. **11**, 82 (2005); gr-qc/0503055.
12. I. Olasagasti and A. Vilenkin, Phys. Rev. D **62**, 044014 (2000); hep-th/0003300.
13. E. Roessl and M. Shaposhnikov, Phys. Rev. D **66**, 084008 (2002); hep-th/0205320.
14. K. A. Bronnikov and B. E. Meierovich, Zh. Éksp. Teor. Fiz. **124**, 5 (2003) [JETP **97**, 1 (2003)]; gr-qc/0301084.
15. B. E. Meierovich, Usp. Fiz. Nauk **171**, 1033 (2001) [Phys. Usp. **44**, 981 (2001)].
16. K. A. Bronnikov, B. E. Meierovich, and E. R. Podolyak, Zh. Éksp. Teor. Fiz. **122**, 459 (2002) [JETP **95**, 392 (2002)].
17. I. Cho and A. Vilenkin, gr-qc/9708005.
18. K. Benson and I. Cho, hep-th/0104067.
19. B. Bajc and G. Gabadadze, Phys. Lett. B **474**, 282 (2000).
20. I. Oda, Phys. Rev. D **62**, 126009 (2000).
21. M. Gogberashvili and P. Midodashvili, Phys. Lett. B **515**, 447 (2001); Europhys. Lett. **61**, 308 (2003); M. Gogberashvili and D. Singleton, hep-th/0305241; I. Oda, Phys. Lett. B **571**, 235 (2003).
22. M. Gogberashvili and D. Singleton, Phys. Lett. B **582**, 95 (2004).

Journal of Experimental and Theoretical Physics, Vol. 101, No. 6, 2005, pp. 1053–1070.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 6, 2005, pp. 1201–1219.

NUCLEI, PARTICLES, FIELDS, GRAVITATION, AND ASTROPHYSICS

Study of the Process $e^+e^- \longrightarrow \pi^+\pi^-$ in the Energy Region $400 < \sqrt{s} < 1000$ MeV¶

M. N. Achasov, K. I. Beloborodov, A. V. Berdyugin, A. G. Bogdanchikov, A. V. Bozhenok, A. D. Bukin, D. A. Bukin, T. V. Dimova, V. P. Druzhinin, V. B. Golubev, I. A. Koop, A. A. Korol, S. V. Koshuba, A. P. Lysenko, A. V. Otboev, E. V. Pakhtusova, S. I. Serednyakov, Yu. M. Shatunov, V. A. Sidorov, Z. K. Silagadze, A. N. Skrinsky, Yu. A. Tikhonov, and A. V. Vasiljev

Budker Institute of Nuclear Physics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia
Novosibirsk State University, Novosibirsk, 630090 Russia
e-mail: achasov@inp.nsk.su
Received June 24, 2005

Abstract—The cross section of the process $e^+e^- \longrightarrow \pi^+\pi^-$ was measured in the spherical neutral detector experiment at the VEPP-2M collider in the energy region $400 < \sqrt{s} < 1000$ MeV. This measurement was based on about 12.4×10^6 selected collinear events, which include 7.4×10^6 $e^+e^- \longrightarrow e^+e^-$, 4.5×10^6 $e^+e^- \longrightarrow \pi^+\pi^-$, and 0.5×10^6 $e^+e^- \longrightarrow \mu^+\mu^-$ selected events. The systematic uncertainty of cross section determination is 1.3%. The ρ-meson parameters were determined as $m_\rho = 774.9 \pm 0.4 \pm 0.5$ MeV, $\Gamma_\rho = 146.5 \pm 0.8 \pm 1.5$ MeV, and $\sigma(\rho \longrightarrow \pi^+\pi^-) = 1220 \pm 7 \pm 16$ nb and the parameters of the *G*-parity suppressed decay $\omega \longrightarrow \pi^+\pi^-$ as $\sigma(\omega \longrightarrow \pi^+\pi^-) = 29.9 \pm 1.4 \pm 1.0$ nb and $\phi_{\rho\omega} = 113.5 \pm 1.3 \pm 1.7°$. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The cross section of the $e^+e^- \longrightarrow \pi^+\pi^-$ process in the energy region $\sqrt{s} < 1000$ MeV can be described within the vector meson dominance model and is determined by the transitions $V \longrightarrow \pi^+\pi^-$ of the light vector mesons ($V = \rho, \omega, \rho', \rho''$) into the final state. The main contribution in this energy region comes from the $\rho \longrightarrow \pi^+\pi^-$ and from the *G*-parity violating $\omega \longrightarrow \pi^+\pi^-$ transitions. Studies of the $e^+e^- \longrightarrow \pi^+\pi^-$ reaction allow determination of the ρ and ω meson parameters and provide information on the *G*-parity violation mechanism.

At low energies, the $e^+e^- \longrightarrow \pi^+\pi^-$ cross section makes the dominant contribution to the well-known ratio

$$R(s) = \frac{\sigma(e^+e^- \longrightarrow \text{hadrons})}{\sigma(e^+e^- \longrightarrow \mu^+\mu^-)},$$

which is used for calculation of the dispersion integrals (for example, for evaluation of the electromagnetic running coupling constant at the *Z*-boson mass $\alpha_{em}(s = m_Z^2)$ or for determination of the hadronic contribution a_μ^{hadr} to the anomalous magnetic moment of the muon, which is nowadays measured with very high accuracy 5×10^{-6}) [1, 2].

Assuming conservation of the vector current (CVC) in the isospin symmetry limit, the spectral function of the $\tau^\pm \longrightarrow \pi^\pm\pi^0\nu_\tau$ decay can be related to the isovector part of the $e^+e^- \longrightarrow \pi^+\pi^-$ cross section. The spectral function was determined with high precision in [3–5]. The comparison of the $e^+e^- \longrightarrow \pi^+\pi^-$ cross section with what follows from the spectral function provides an accurate test of the CVC hypothesis.

The process $e^+e^- \longrightarrow \pi^+\pi^-$ in the energy region $\sqrt{s} < 1000$ MeV has been studied in several experiments [6–19] for more than 30 years. In the present work, we report on the results of the $e^+e^- \longrightarrow \pi^+\pi^-$ cross section measurement with a spherical neutral detector (SND) at $390 \leq \sqrt{s} \leq 980$ MeV.

2. EXPERIMENT

The SND [20] operated from 1995 to 2000 at the VEPP-2M [21] collider; it worked in the energy range $\sqrt{s}$ from 360 to 1400 MeV. The detector contains several subsystems. The tracking system includes two cylindrical drift chambers. The three-layer spherical electromagnetic calorimeter is based on NaI(Tl) crystals. The muon–veto system consists of plastic scintillation counters and two layers of streamer tubes. The

¶ The text was submitted by the authors in English.

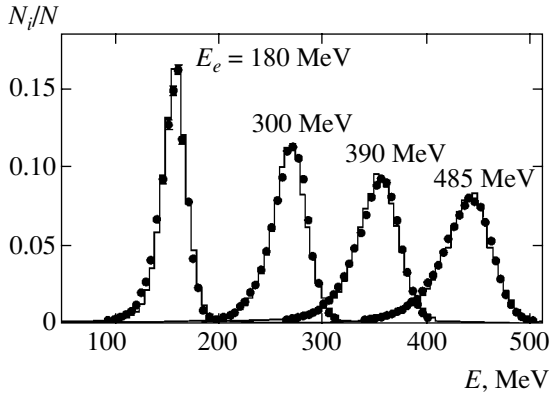


Fig. 1. Energy deposition spectra for electrons with energies of 180, 300, 390, and 485 MeV in experiment (dots) and MC simulation (histogram).

calorimeter energy and angular resolutions depend on the photon energy as

$$(\sigma_E/E)\% = 4.2\%/\sqrt[4]{E[\text{GeV}]}$$

and

$$\sigma_{\phi,\theta} = 0.82^\circ/\sqrt{E[\text{GeV}]} \oplus 0.63^\circ.$$

The tracking system angular resolution is about 0.5° and 2° for the azimuthal and polar angles, respectively.

In 1996–2000, the SND collected data in the energy region $\sqrt{s} < 980$ MeV with an integrated luminosity of about 10.0 pb^{-1}. The beam energy was calculated from the magnetic field value in the bending magnets of the collider. The accuracy of the energy setting is about 0.1 MeV. The beam energy spread varies in the range from 0.06 MeV at $\sqrt{s}$ = 360 MeV to 0.35 MeV at $\sqrt{s}$ = 970 MeV.

3. DATA ANALYSIS

The cross section of the $e^+e^- \longrightarrow \pi^+\pi^-$ process was measured as follows.

(1) The collinear events $e^+e^- \longrightarrow e^+e^-$, $\pi^+\pi^-$, $\mu^+\mu^-$ were selected.

(2) The selected events were sorted into the two classes: e^+e^- and $\pi^+\pi^-$, $\mu^+\mu^-$ using the energy deposition in the calorimeter layers.

(3) The $e^+e^- \longrightarrow e^+e^-$ events were used for integrated luminosity determination. The events of the $e^+e^- \longrightarrow \mu^+\mu^-$ process were subtracted according to the theoretical cross section, integrated luminosity, and detection efficiency.

(4) To determine the cross section of the $e^+e^- \longrightarrow \pi^+\pi^-$ process, the number of $e^+e^- \longrightarrow \pi^+\pi^-$ events in each energy point were normalized to the integrated luminosity and divided by the detection efficiency and radiative correction.

The detection efficiency was obtained from Monte Carlo (MC) simulation [20]. The MC simulation of SND is based on the UNIMOD [22] package. The SND geometrical model description comprises about 10000 distinct volumes and includes details of the SND's design. The primary generated particles are tracked through the detector media with the following effects taken into account: ionization losses, multiple scattering, bremsstrahlung of electrons and positrons, Compton effect and Rayleigh scattering, e^+e^- pair production by photons, photoeffect, unstable particles decay, interaction of stopped particles, and nuclear interaction of hadrons [23–25]. After that, the signals produced in each detector element are simulated. The electronic noise, signal pile up, actual time, and amplitude resolutions of the electronics channels and broken channels were taken into account during processing of the Monte Carlo events to provide an adaptable account of variable experimental conditions.

The Monte Carlo simulation of the processes $e^+e^- \longrightarrow e^+e^-$, $\mu^+\mu^-$, $\pi^+\pi^-$ was based on the formula obtained in [26–28]. The simulation of the process $e^+e^- \longrightarrow e^+e^-$ was performed with the cut $30^\circ < \theta_{e^\pm} < 150^\circ$ on the polar angles of the final electron and positron.

The $e^+e^- \longrightarrow e^+e^-$, $\mu^+\mu^-$, and $\pi^+\pi^-$ events differ by energy deposition in the calorimeter. In the $e^+e^- \longrightarrow e^+e^-$ events, the electrons produce an electromagnetic shower with the most probable energy losses about 0.92 of the initial particle energy. The distributions of the energy deposition of the electrons with different energies are shown in Fig. 1. The experimental and simulated spectra are in good agreement. Muons lose their energy by ionization of the calorimeter material through which they pass, and their energy deposition spectra are well modeled in simulation (Fig. 2). Similar ionization losses are experienced by charged pions, and this part of the charged pion energy deposition is well described by simulation (Fig. 3). However, pions also lose their energy due to nuclear interactions, which is not so accurately reproduced in simulation. This leads to some difference in the energy deposition spectra in experiment and simulation for charged pions (Fig. 4).

The discrimination between electrons and pions in the SND is based on the difference in the longitudinal energy deposition profiles (deposition in calorimeter layers) for these particles. To fully use the correlations between energy depositions in the calorimeter layers, the corresponding separation parameter was based on the neural network approach [29]. For each energy point, the neural network, a multilayer perceptron, was constructed. The network had an input layer consisting of seven neurons, two hidden layers with 20 neurons each, and an output layer with one neuron. As the input data, the network used the energy depositions of parti-

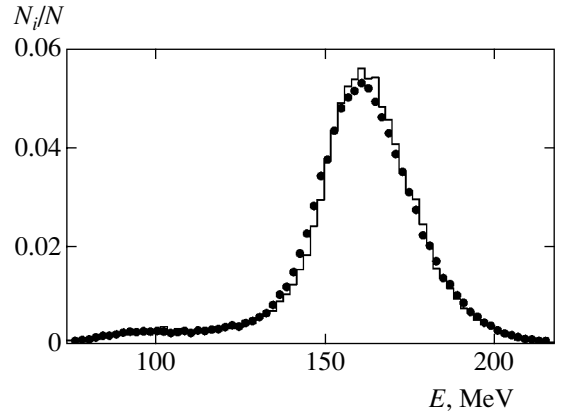


Fig. 2. Energy deposition spectra for the 500 MeV muons in experiment (dots) and MC simulation (histogram).

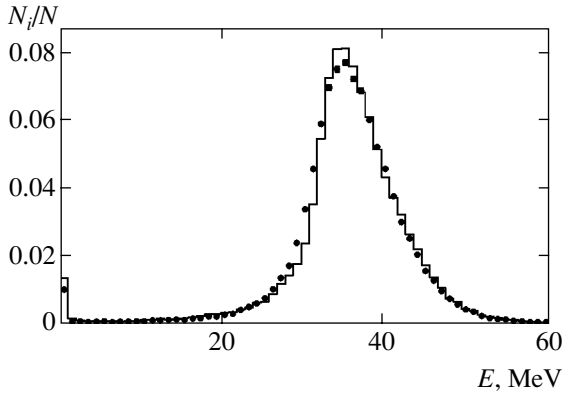


Fig. 3. Spectra of the ionization losses of the pions with the energy $E_\pi > 360$ MeV in the first calorimeter layer. Dots, experiment; histogram, MC simulation.

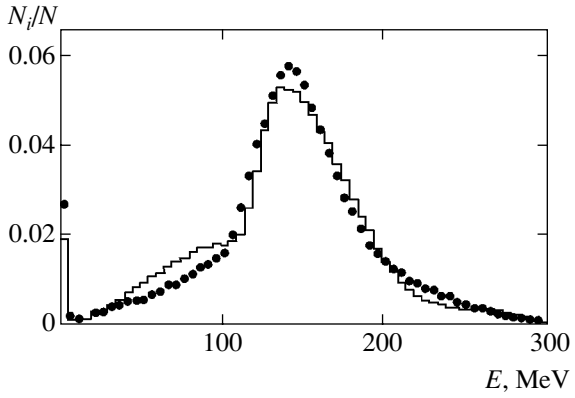


Fig. 4. Energy deposition spectra of the pions with the energy $E_\pi = 300$ MeV. Dots, experiment; histogram, MC simulation.

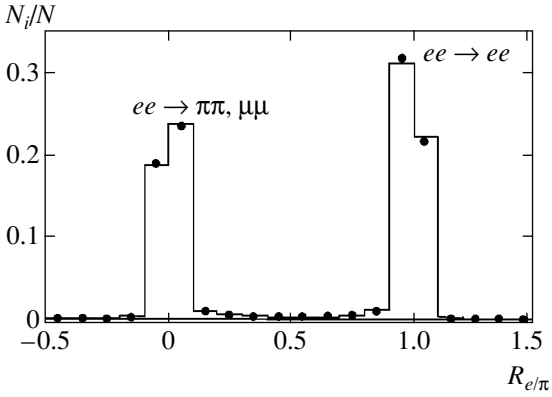


Fig. 5. The e/π discrimination parameter distribution for all collinear events in the energy region $\sqrt{s}$ from 880 to 630 MeV. Dots, experiment; histogram, MC simulation.

cles in calorimeter layers and the polar angle of one of the particles. The output signal $R_{e/\pi}$ is a number in the interval from –0.5 to 1.5. The network was trained by using simulated $e^+e^- \longrightarrow \pi^+\pi^-$ and $e^+e^- \longrightarrow e^+e^-$ events. The distribution of the discrimination parameter $R_{e/\pi}$ is shown in Fig. 5. The $e^+e^- \longrightarrow e^+e^-$ events are located in the region $R_{e/\pi} > 0.5$, and the $e^+e^- \longrightarrow \pi^+\pi^-$, $\mu^+\mu^-$ events are located at $R_{e/\pi} < 0.5$.

3.1. Selection Criteria

During the experimental runs, the first-level trigger [20] selects events with one or more tracks in the tracking system and with two clusters in the calorimeter with a spatial angle between the clusters of more than 100°. The threshold of energy deposition in a cluster was equal to 25 MeV. The threshold of the total energy deposition in the calorimeter was set equal to 140 MeV in the energy region $\sqrt{s} \geq 850$ MeV, and to 100 MeV, or was absent altogether, below 850 MeV. During processing of the experimental data, event reconstruction is performed [20, 30]. For further analysis, events containing two charged particles with $|z| < 10$ cm and $r <$ 1 cm were selected. Here, z is the coordinate of the charged particle production point along the beam axis (the longitudinal size of the interaction region depends on the beam energy and varies from 1.5 to 2.5 cm) and r is the distance between the charged particle track and the beam axis in the $r\phi$ plane. The polar angles of the charged particles were bounded by the criterion $55° < \theta < 125°$, and the energy deposition of each of them was required to be greater than 50 MeV. The following cuts on the acollinearity angles in the azimuthal and polar planes were applied: $|\Delta\phi| < 10°$ and $|\Delta\theta| < 10°$. In the event sample selected under these conditions, one has the $e^+e^- \longrightarrow e^+e^-$, $\pi^+\pi^-$, $\mu^+\mu^-$ events, cosmic muons background, and a small contribution from the $e^+e^- \longrightarrow \pi^+\pi^-\pi^0$ reaction at $\sqrt{s} \approx m_\omega$. The muon system veto was used for suppressing the cosmic muon background (veto = 0).

3.2. The Background from Cosmic Muons and from the $e^+e^- \longrightarrow \pi^+\pi^-\pi^0$ Process

The number of background events from the $e^+e^- \longrightarrow \pi^+\pi^-\pi^0$ process was estimated as

$$N_{3\pi}(s) = \sigma_{3\pi}(s)\epsilon_{3\pi}(s)IL(s), \quad (1)$$

where $\sigma_{3\pi}(s)$ is the cross section of the $e^+e^- \longrightarrow \pi^+\pi^-\pi^0$ process with the radiative corrections taken into account, $IL(s)$ is the integrated luminosity, and $\epsilon_{3\pi}(s)$ is the detection probability for the background process obtained from the simulation under the selection criteria described above. The values of $\sigma_{3\pi}(s)$ were taken from the SND measurements [31]. Although $\sigma_{3\pi}(m_\omega) \approx$ 1300 nb, the $e^+e^- \longrightarrow 3\pi$ process contribution to the total number of collinear events at the ω resonance peak is less than 0.3%. The leading role in the suppression of this background was played by the cuts on the acollinearity angles $\Delta\theta$ and $\Delta\phi$. In order to check the estimate in (1), the events containing two or more photons with energy depositions of more than 200 MeV were considered.

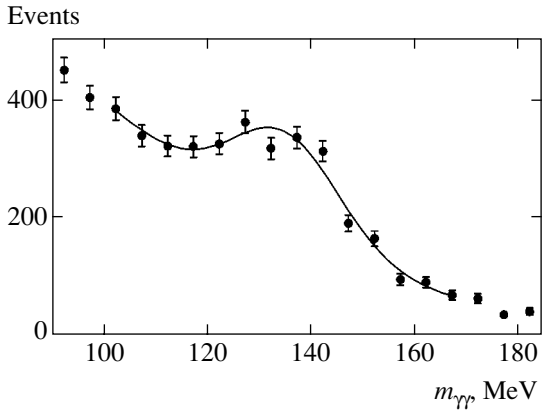


Fig. 6. Two-photon invariant mass $m_{\gamma\gamma}$ distribution at $\sqrt{s} \approx m_\omega$.

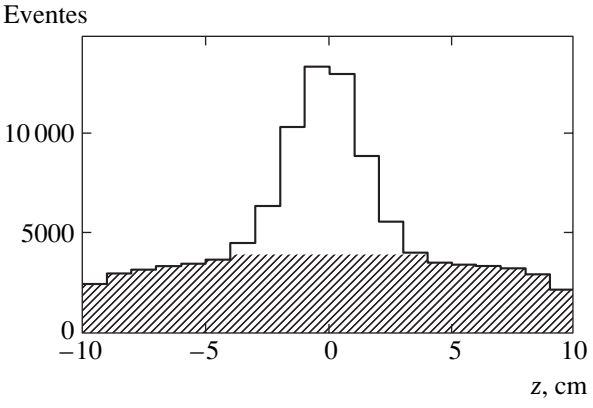


Fig. 7. Distribution of the z coordinate of the charged particle production point along the beam axis for collinear events at $\sqrt{s}$ = 180 MeV. Histogram, all events; dashed distribution, events with muon system veto (veto = 1).

Because our selection criteria select the $e^+e^- \longrightarrow 3\pi$ events with collinear charged pions and, therefore, the neutral pion in these events has relatively low energy, the constraint on the photon energy deposition greatly suppresses the events other than $e^+e^- \longrightarrow 3\pi$ events. To obtain the $e^+e^- \longrightarrow 3\pi$ event number $n_{3\pi}$, the invariant mass spectrum $m_{\gamma\gamma}$ (Fig. 6) was fitted by the sum of a Gaussian function and a second-order polynomial:

$$G(m_{\gamma\gamma})n_{3\pi} + P_2(m_{\gamma\gamma})(n - n_{3\pi}).$$

The value of $n_{3\pi}$ agrees with events number calculated according to (1).

The cosmic muon background was suppressed by the muon–veto system. The z coordinate distribution for the charged particle production point along the beam axis is shown in Fig. 7 for collinear events. The e^+e^- annihilation events have the Gaussian distribution peaked at $z = 0$, while the cosmic background distribution is nearly uniform and clearly extends outside the peak. As Fig. 7 shows, the muon system veto (veto = 1) separates cosmic muons from the e^+e^- annihilation events. The residual event number of the cosmic muon background was estimated from the formula

$$N_\mu = \nu_\mu T, \quad (2)$$

where $\nu_\mu \approx 1.3 \times 10^{-3}$ Hz is the frequency of cosmic background detection under the applied selection criteria and T is the time the data was taken. The value of ν_μ was obtained by using data collected in special runs without beams in the collider. The first-level trigger counting rate in these runs was 2 Hz. The contribution of the cosmic background to the total number of selected collinear events depends on the energy $\sqrt{s}$ and varies from 0.1 to 1%.

The $e^+e^- \longrightarrow \pi^+\pi^-\pi^0$ events are concentrated in the $R_{e/\pi}$ discrimination parameter region $R_{e/\pi} < 0.5$. The cosmic background events at energies of $\sqrt{s} > 600$ MeV also fall in the area $R_{e/\pi} < 0.5$, because the energy deposition of the cosmic muons is much lower than the energy deposition in the $e^+e^- \longrightarrow e^+e^-$ events. For lower center-of-mass energies, the cosmic background moves to the area $R_{e/\pi} > 0.5$, because the energy depositions are close in this case.

3.3. Detection Efficiency

The $\Delta\phi$ and $\Delta\theta$ distributions of the $e^+e^- \longrightarrow e^+e^-$ and $e^+e^- \longrightarrow \pi^+\pi^-$ events are shown in Figs. 8–11. Experiment and simulation agree rather well. As a measure of

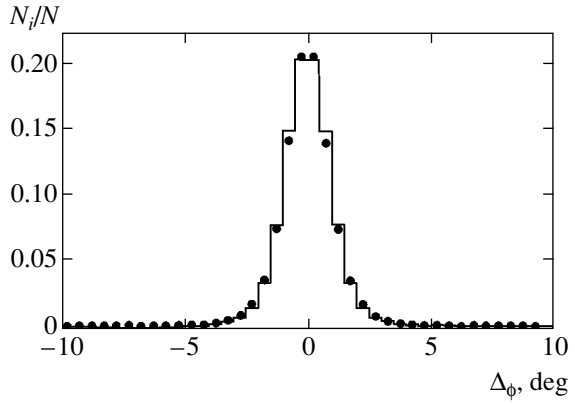


Fig. 8. The $\Delta\phi$ distribution of the $e^+e^- \longrightarrow e^+e^-$ events. Dots, experiment; histogram, MC simulation.

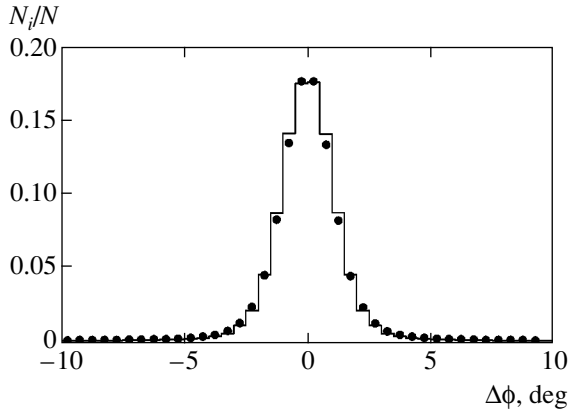


Fig. 9. The $\Delta\phi$ distribution of the $e^+e^- \longrightarrow \pi^+\pi^-$ events. Dots, experiment; histogram, MC simulation.

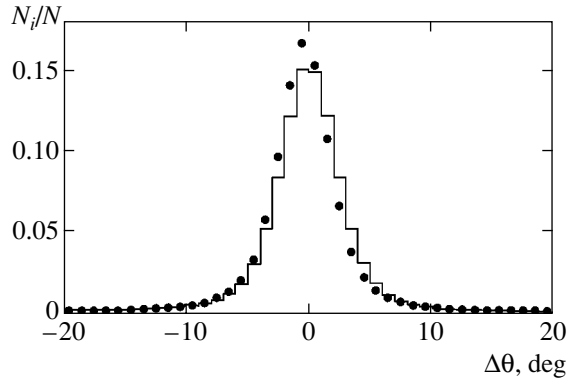


Fig. 10. The $\Delta\theta$ distribution of the $e^+e^- \longrightarrow e^+e^-$ events. Dots, experiment; histogram, MC simulation.

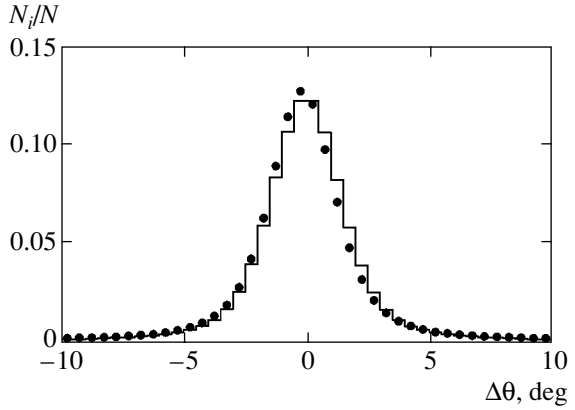


Fig. 11. The $\Delta\theta$ distribution of the $e^+e^- \longrightarrow \pi^+\pi^-$ events. Dots, experiment; histogram, MC simulation.

the systematic uncertainty due to the $\Delta\theta$ cut, the following value was used:

$$\delta_{\Delta\theta} = \frac{\delta^{\pi\pi}_{\Delta\theta}}{\delta^{ee}_{\Delta\theta}}, \tag{3}$$

where

$$\delta^x_{\Delta\theta} = \frac{n_x(|\Delta\theta| < 10^\circ)}{N_x(|\Delta\theta| < 20^\circ)} \Big/ \frac{m_x(|\Delta\theta| < 10^\circ)}{M_x(|\Delta\theta| < 20^\circ)},$$
$$x = \pi\pi(ee).$$

Here, $n_x(|\Delta\theta| < 10^\circ)$ and $m_x(|\Delta\theta| < 10^\circ)$ are the numbers of experimental and simulated events selected under the condition $|\Delta\theta| < 10^\circ$, while $N_x(|\Delta\theta| < 20^\circ)$ and $M_x(|\Delta\theta| < 20^\circ)$ are the numbers of experimental and simulated events with $|\Delta\theta| < 20^\circ$. $\delta_{\Delta\theta}$ does not depend on energy, its average value is equal to 0.999, and it has a systematic spread of 0.4%. This systematic spread was added to the error of the cross section measurement at each energy point. The systematic error due to the $\Delta\phi$ cut is significantly lower and was neglected.

The polar angle distributions for the $e^+e^- \longrightarrow e^+e^-$ and $e^+e^- \longrightarrow \pi^+\pi^-$ processes are shown in Figs. 12 and 13. The ratio of these θ distributions is shown in Fig. 14. The experimental and simulated distributions are in agreement. To estimate the systematic inaccuracy due to the θ angle selection cut, the following ratio was used:

$$\delta_\theta = \frac{\delta(\theta_x)}{\delta(55^\circ)}, \tag{4}$$

where

$$\delta(\theta_x) = \frac{N_{\pi\pi}(\theta_x < \theta < 180^\circ - \theta_x)}{N_{ee}(\theta_x < \theta < 180^\circ - \theta_x)} \Big/ \frac{M_{\pi\pi}(\theta_x < \theta < 180^\circ - \theta_x)}{M_{ee}(\theta_x < \theta < 180^\circ - \theta_x)},$$
$$50^\circ < \theta_x < 90^\circ.$$

Here, $N_{\pi\pi}(\theta_x < \theta < 180^\circ - \theta_x)$, $N_{ee}(\theta_x < \theta < 180^\circ - \theta_x)$, $M_{\pi\pi}(\theta_x < \theta < 180^\circ - \theta_x)$, and $M_{ee}(\theta_x < \theta < 180^\circ - \theta_x)$ are the experimental and simulated $e^+e^- \longrightarrow \pi^+\pi^-$ and

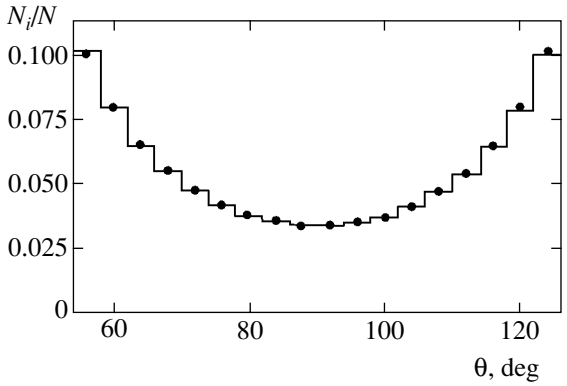


Fig. 12. The θ angle distribution of the $e^+e^- \longrightarrow e^+e^-$ events. Dots, experiment; histogram, MC simulation.

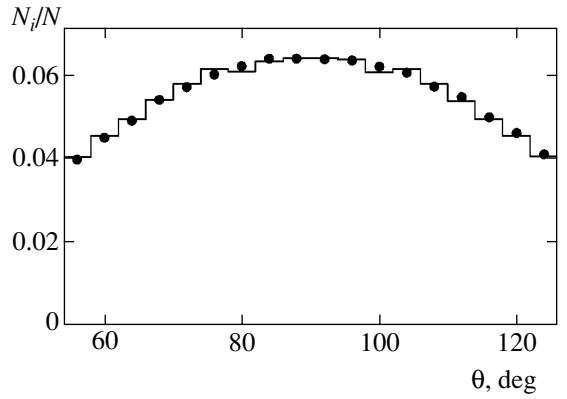


Fig. 13. The θ angle distribution of the $e^+e^- \longrightarrow \pi^+\pi^-$ events. Dots, experiment; histogram, MC simulation.

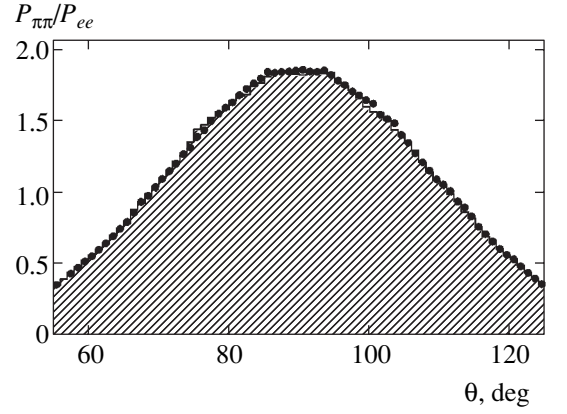


Fig. 14. The ratio of θ distributions of the $e^+e^- \longrightarrow \pi^+\pi^-$ and $e^+e^- \longrightarrow e^+e^-$ processes. Dots, experiment; histogram, MC simulation.

$e^+e^- \longrightarrow e^+e^-$ event numbers in the angular range $\theta_x < \theta < 180° - \theta_x$. The maximal difference of δ_θ from unity was found to be 0.8%. This value was taken as a systematic error $\delta_\theta = 0.8\%$ associated with the angular selection cut.

In the tracking system, the particle track can be lost due to reconstruction inefficiency. The probabilities of finding the track were determined by using the experimental data themselves. It was found to be $\varepsilon_e \approx 0.996$ for electrons and $\varepsilon_\pi \approx 0.995$ for pions. In simulations, these values do not actually differ from unity, while in reality, the probability of finding the track for electrons is slightly greater than for pions. Therefore, the detection efficiency was multiplied by the correction coefficient

$$\delta_{\text{rec}} = \left(\frac{\varepsilon_\pi}{\varepsilon_e}\right)^2 = 0.997. \quad (5)$$

Pions can be lost due to the nuclear interaction in the detector material before the tracking system, for example, via the reaction $\pi^\pm N \longrightarrow \pi^\pm N$, with the final pion scattered at a large angle, or via the charge exchange reaction $\pi^\pm N \longrightarrow \pi^0 N$. As a measure of systematic inaccuracy associated with this effect, the difference from unity of the following quantity was used:

$$\delta_{\text{nucl}} = \left(\frac{1 - \dfrac{n}{3N}}{1 - \dfrac{m}{3M}}\right)^2. \quad (6)$$

Here, N and M are the pion numbers in experiment and simulation; n and m are the pion numbers in experiment and simulation in the case where a track in the drift chamber nearest to the beam pipe was detected, but the corresponding track in the second drift chamber and associated cluster in the calorimeter was not found. The probability of particle loss was divided by 3, the ratio of the amounts of matter between the drift chambers and before the tracking system. The deviation of δ_{nucl} from 1 was taken as a systematic error, $\sigma_{\text{nucl}} = 0.2\%$.

Uncertainties in simulation of pion nuclear interactions imply that the cut on the particle energy deposition leads to an inaccuracy in the detection efficiency of the $e^+e^- \longrightarrow \pi^+\pi^-$ process. To take this inaccuracy into account, the detection efficiency was multiplied by the correction coefficients. The correction coefficients were obtained by using events of the $e^+e^- \longrightarrow \pi^+\pi^-\pi^0$ reaction [30–32]. Pion energies in the $e^+e^- \longrightarrow \pi^+\pi^-\pi^0$ events were determined via the kinematic fit. The pion energies were divided into 10-MeV-wide bins. For each bin, the correction coefficient (Fig. 15) was obtained as

$$\delta_{E>50} = \left[\frac{n_i/N_i}{m_i/M_i}\right]^2, \quad (7)$$

where i is the bin number, N_i and M_i are the pion numbers in experiment and simulation selected in the ith bin by the kinematic fit without any cut on the energy deposition in the calorimeter, and n_i and m_i are the pion

numbers in experiment and simulation under the condition that the pion energy deposition is greater than 50 MeV. To estimate systematic errors in determining these correction coefficients, we considered the ratio of the probability that both pions in simulated $e^+e^- \longrightarrow \pi^+\pi^-$ events have an energy deposition of more than 50 MeV to the quantity $(m_i/M_i)^2$. This ratio is 0.994 at $\sqrt{s} > 420$ MeV and about 0.97 at $\sqrt{s} < 420$ MeV. The difference of this ratio from unity was taken as a systematic error $\sigma_{E>50}$ in determining the $\delta_{E>50}$ correction coefficient: $\sigma_{E>50} = 0.6\%$ at $\sqrt{s} > 420$ MeV and $\sigma_{E>50} = 3\%$ at $\sqrt{s} < 420$ MeV.

In the energy region $\sqrt{s} = 840$–970 MeV, the probability of hitting the muon–veto system for muons and pions varies from 1 to 93% and from 0.5 to 3%, respectively. The usage of the muon system veto for event selection (veto = 0) leads to inaccuracy in the determination of the measured cross section due to the uncertainty in the simulation of the muons and pions traversing the detector at $\sqrt{s} > 840$ MeV. To obtain the necessary corrections, the events close to the median plane $\phi < 10°$, $170° < \phi < 190°$, $\phi > 350°$, where the cosmic background is minimal, were used. The $e^+e^- \longrightarrow \pi^+\pi^-$ cross section was measured with (veto = 0) and without (veto $\geq$ 0) using the muon system, and the following correction coefficient was obtained for each energy point:

$$\delta_{\text{veto}} = \frac{\sigma(e^+e^- \longrightarrow \pi^+\pi^-;\ \text{veto} \geq 0)}{\sigma(e^+e^- \longrightarrow \pi^+\pi^-;\ \text{veto} = 0)}. \qquad (8)$$

It was found that $\delta_{\text{veto}} = 0.95$ at $\sqrt{s} = 970$ MeV and quickly rises to 1 for lower energies.

The detection efficiencies of the processes $e^+e^- \longrightarrow \pi^+\pi^-$, $\mu^+\mu^-$, and e^+e^- after all the applied corrections are shown in Fig. 16. The detection efficiency is independent of energy for the $e^+e^- \longrightarrow e^+e^-$ reaction, but depends on it for the $e^+e^- \longrightarrow \mu^+\mu^-$ and $\pi^+\pi^-$ processes. The decrease in the $e^+e^- \longrightarrow \mu^+\mu^-$ process detection efficiency at $\sqrt{s} > 800$ MeV is caused by the fact that the probability of muons hitting the muon system increases with energy. The detection efficiency of the $e^+e^- \longrightarrow \pi^+\pi^-$ process at $\sqrt{s} > 500$ MeV is determined mainly by the cuts on the pion angles. Below 500 MeV, the detection efficiency decreases due to the cut on the pion energy deposition in the calorimeter. A statistical error of ≤1% for the detection efficiency determination was added to the cross section measurement error at each energy point. The total systematic error of determining the detection efficiency $\sigma_{\text{eff}} = \sigma_{E>50} \oplus \sigma_{\text{nucl}} \oplus \sigma_\theta$ is $\sigma_{\text{eff}} = 1\%$ at $\sqrt{s} \geq 420$ MeV and $\sigma_{\text{eff}} = 3.1\%$ at $\sqrt{s} < 420$ MeV.

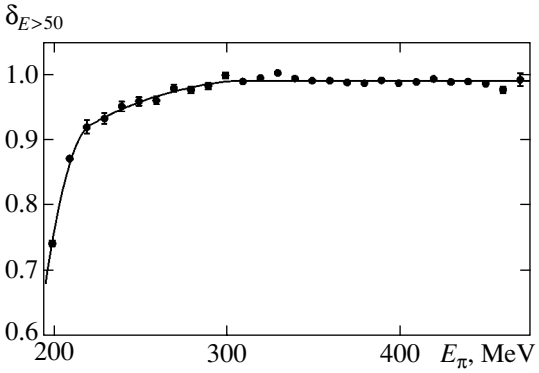


Fig. 15. The $d_{E>50}$ correction coefficient associated with the pion energy deposition cut vs. the pion energy E_π.

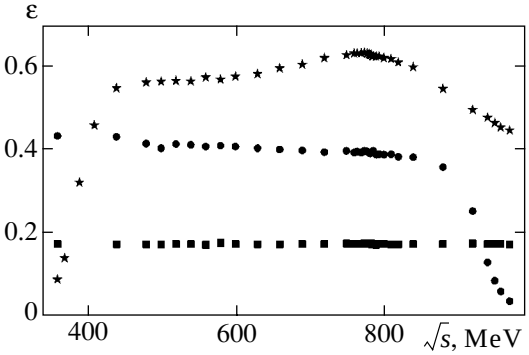


Fig. 16. The detection efficiencies $\varepsilon_{\pi\pi}$ (★), ε_{ee} (■), and $\varepsilon_{\mu\mu}$ (●) of the $e^+e^- \longrightarrow \pi^+\pi^-$, $\mu^+\mu^-$, and e^+e^- processes.

3.4. Measurement of the $e^+e^- \longrightarrow \pi^+\pi^-$ Cross Section

The number of selected events in the regions $R_{e/\pi} < 0.5$ and $R_{e/\pi} > 0.5$ are

$$N = N_{\pi\pi} + N_{ee} + N_{\mu\mu} + N_\mu + N_{3\pi}, \qquad (9)$$

$$M = M_{\pi\pi} + M_{ee} + M_{\mu\mu} + M_\mu + M_{3\pi}. \qquad (10)$$

Here, N and M are the event numbers in the respective regions $R_{e/\pi} < 0.5$ and $R_{e/\pi} > 0.5$. N_μ, M_μ and $N_{3\pi}$, $M_{3\pi}$ are the numbers of background events due to cosmic muons and the $e^+e^- \longrightarrow \pi^+\pi^-\pi^0$ process, calculated as described above. The $e^+e^- \longrightarrow \mu^+\mu^-$ process event number can be written as

$$N_{\mu\mu} = \sigma_{\mu\mu}\varepsilon_{\mu\mu}(1 - \epsilon_{\mu\mu})IL, \qquad (11)$$

$$M_{\mu\mu} = \sigma_{\mu\mu}\varepsilon_{\mu\mu}\epsilon_{\mu\mu}IL, \qquad (12)$$

where $\sigma_{\mu\mu}$ is the $e^+e^- \longrightarrow \mu^+\mu^-$ process cross section obtained according to [27], $\varepsilon_{\mu\mu}$ is the process detection

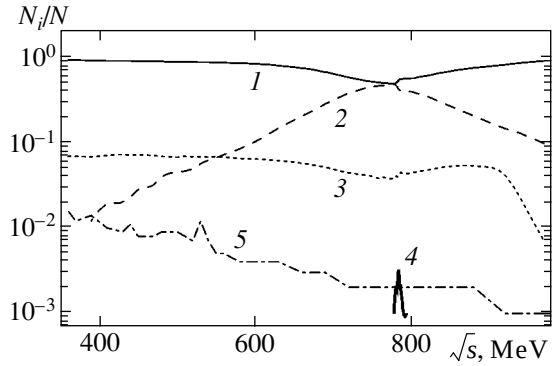


Fig. 17. The percentage of the $e^+e^- \longrightarrow e^+e^-$ (*1*), $\pi^+\pi^-$ (*2*), $\mu^+\mu^-$ (*3*), $\pi^+\pi^-\pi^0$ (*4*), and cosmic background (*5*) vs. the energy $\sqrt{s}$.

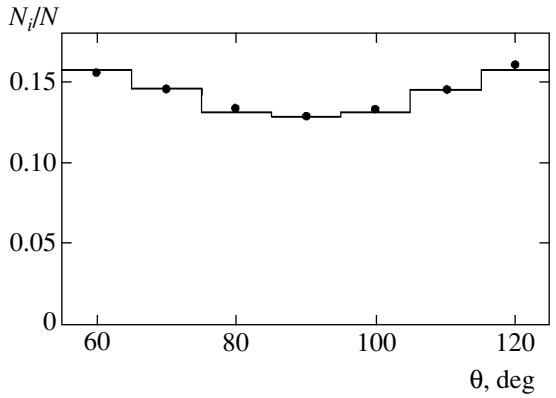


Fig. 18. The θ angle distributions of all collinear events at $\sqrt{s}$ from 880 to 630 MeV. Dots, experiment; histogram, MC simulation.

efficiency, $\epsilon_{\mu\mu}$ is the probability for the $e^+e^- \longrightarrow \mu^+\mu^-$ process events to have $R_{e/\pi} > 0.5$, and IL is the integrated luminosity,

$$IL = \frac{M_{ee}}{\sigma_{ee}\varepsilon_{ee}\epsilon_{ee}}, \tag{13}$$

where ε_{ee} and ϵ_{ee} are the detection efficiency and the probability of having $R_{e/\pi} > 0.5$ for the process $e^+e^- \longrightarrow e^+e^-$, and σ_{ee} is the process cross section with the $30° < \theta < 150°$ angular cut for the electron and positron in the final state. The cross section σ_{ee} was calculated using the BHWIDE 1.04 [33] code with the accuracy 0.5%. The $e^+e^- \longrightarrow \pi^+\pi^-$ process event number with $R_{e/\pi} > 0.5$ and the $e^+e^- \longrightarrow e^+e^-$ process event number with $R_{e/\pi} < 0.5$ can be written as

$$N_{ee} = \frac{1-\epsilon_{ee}}{\epsilon_{ee}} M_{ee} = \lambda_{ee} M_{ee},$$

$$M_{\pi\pi} = \frac{1-\epsilon_{ee}}{\epsilon_{ee}} N_{\pi\pi} = \lambda_{\pi\pi} N_{\pi\pi}.$$

The $e^+e^- \longrightarrow e^+e^-$ process event number with $R_{e/\pi} > 0.5$ and the $e^+e^- \longrightarrow \pi^+\pi^-$ process event number with $R_{e/\pi} < 0.5$ are

$$M_{ee} = \frac{M - M_\mu - \lambda_{\pi\pi}(N - N_\mu)}{\kappa - \Delta\lambda_{\pi\pi}}, \tag{14}$$

$$N_{\pi\pi} = N - N_\mu - M_{ee}\Delta, \tag{15}$$

where

$$\Delta = \lambda_{ee} + \frac{\sigma_{\mu\mu}\varepsilon_{\mu\mu}(1-\epsilon_{\mu\mu}) + N_{3\pi}/IL}{\sigma_{ee}\varepsilon_{ee}\epsilon_{ee}},$$

$$\kappa = 1 + \frac{\sigma_{\mu\mu}\varepsilon_{\mu\mu}\epsilon_{\mu\mu} + M_{3\pi}/IL}{\sigma_{ee}\varepsilon_{ee}\epsilon_{ee}}.$$

The percentage of each process in the selected events versus the energy $\sqrt{s}$ is shown in Fig. 17. The experimental angular distributions agree with the sum of distributions for each process weighted according to its contribution (Fig. 18).

The $e^+e^- \longrightarrow \pi^+\pi^-$ process cross section is calculated from the formula

$$\sigma_{\pi\pi} = \frac{N_{\pi\pi}}{IL\varepsilon_{\pi\pi}(1-\epsilon_{\pi\pi})} = \frac{\sigma_{ee}\varepsilon_{ee}\epsilon_{ee}}{\varepsilon_{\pi\pi}(1-\epsilon_{\pi\pi})}\left(\frac{\kappa - \Delta\lambda_{\pi\pi}}{\dfrac{M-M_\mu}{N-N_\mu}} - \Delta\right). \tag{16}$$

To estimate the systematic uncertainty due to e–π discrimination, the pseudo $\pi\pi$ and pseudo ee events in the experiment and simulation were formed. The pseudo $\pi\pi$ events were constructed by using pions from the $e^+e^- \longrightarrow \pi^+\pi^-\pi^0$ reaction. To construct the pseudo $\pi\pi$ event with pions having an energy E_0, two charged pions with energies E_π such that $|E_0 - E_\pi| < 10$ MeV were used from two separate $e^+e^- \longrightarrow \pi^+\pi^-\pi^0$ events. Of course, such pseudo $\pi\pi$ events are in general not collinear, but this is irrelevant for our purposes here. The pseudo ee event was constructed analogously from the particles of two separate collinear events such that their partners in these events have energy depositions in the calorimeter layers typical for electrons. Figures 19 and 20 show the probabilities for the discrimination parameter to have values less than some magnitude in experiment and simulation for such pseudo events. Using these distributions, the corrections to the probabilities for the separation parameter $R_{e/\pi}$ to be greater or less than 0.5 was obtained. The difference between cross sections measured with and without these corrections

was taken as a systematic error, and its value does not exceed 0.5% for different energy points.

The cross sections obtained, together with the radiative corrections δ_{rad}, including the initial- and final-state radiation, are presented in Table 1. The δ_{rad} radiative correction was calculated according to [28]. The accuracy of its determination is 0.2%. Given the radiative corrections, the Born cross section for the $e^+e^- \longrightarrow \pi^+\pi^-$ process can be obtained as

$$\sigma_0(s) = \frac{\sigma_{\pi\pi}(s)}{\delta_{rad}(s)}. \quad (17)$$

The value of $\delta_{rad}(s)$ depends on the cross section at lower energies, and it was therefore calculated iteratively. The iteration stops when its value changes by not more than 0.1% in consecutive iterations. The form factor values

$$|F_\pi(s)|^2 = \frac{3s}{\pi\alpha^2\beta^3}\sigma_{\pi\pi}(s), \quad \beta = \sqrt{1-\frac{4m_\pi^2}{s}}$$

are also listed in Table 1. To evaluate the value of

$$R(s) = \frac{\sigma(e^+e^- \longrightarrow \text{hadrons})}{\sigma(e^+e^- \longrightarrow \mu^+\mu^-)},$$

which is used in calculating dispersion integrals, the bare cross section $e^+e^- \longrightarrow \pi^+\pi^-$ is used (the cross section without vacuum polarization contribution but with the final-state radiation taken into account),

$$\sigma_{\pi\pi}^{pol}(s) = \sigma_0(s)|1-\Pi(s)|^2\left(1+\frac{\alpha}{\pi}a(s)\right), \quad (18)$$

where $\Pi(s)$ is the polarization operator calculated according to [27] from the known $e^+e^- \longrightarrow$ hadrons cross section [34]. The last factor takes the final-state radiation into account, and $a(s)$ has the form [35]

$$a(s) = \frac{1+\beta^2}{\beta}\left[4\mathrm{Li}_2\left(\frac{1-\beta}{1+\beta}\right)+2\mathrm{Li}_2\left(-\frac{1-\beta}{1+\beta}\right)\right.$$

$$\left.-3\ln\frac{2}{1+\beta}\ln\frac{1+\beta}{1-\beta}-2\ln\beta\ln\frac{1+\beta}{1-\beta}\right]$$

$$-3\ln\frac{4}{1-\beta^2}-4\ln\beta+\frac{1}{\beta^3}\left[\frac{5}{4}(1+\beta^2)^2-2\right]$$

$$\times\ln\frac{1+\beta}{1-\beta}+\frac{3}{2}\frac{1+\beta^2}{\beta^2}.$$

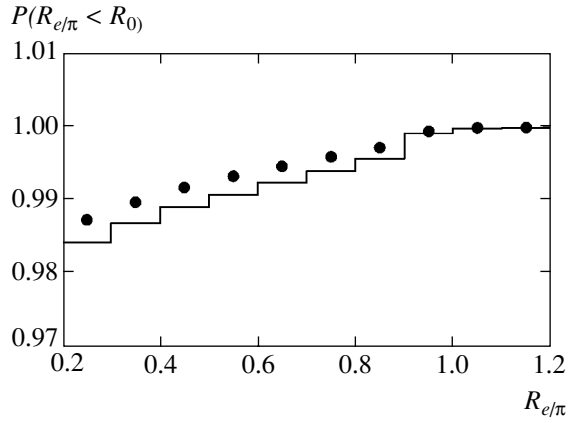


Fig. 19. The probability of the pseudo $\pi\pi$ events to have an $R_{e/\pi}$ value less than some R_0. Dots, experiment; histogram, MC simulation.

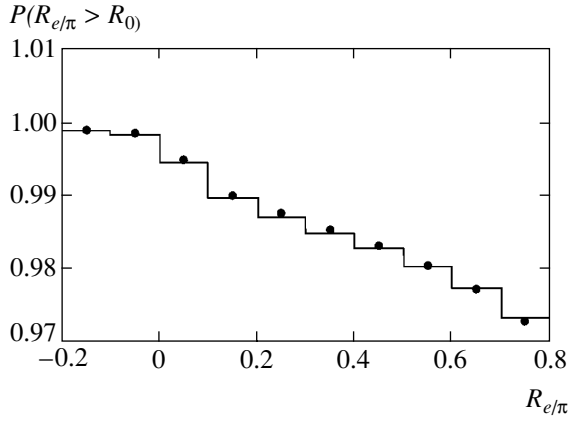


Fig. 20. The probability of the pseudo *ee* events to have an $R_{e/\pi}$ value greater than some R_0. Dots, experiment; histogram, MC simulation.

Here,

$$\mathrm{Li}_2(x) = -\int_0^x dt\ln(1-t)/t.$$

The values of $\sigma_{\pi\pi}^{pol}(s)$ are listed in Table 1.

The total systematic error of the cross section determination is

$$\sigma_{sys} = \sigma_{eff} \oplus \sigma_{sep} \oplus \sigma_{IL} \oplus \sigma_{rad}.$$

Here, σ_{eff} is the systematic error of the detection efficiency determination, σ_{sep} is the systematic error associated with the e–π separation, σ_{IL} is the systematic error of the integrated luminosity determination, and σ_{rad} is the uncertainty of the radiative correction calculation. The magnitudes of various contributions to the total systematic error are shown in Table 2. The total systematic error of the cross section determinations is

Table 1. The results of the $e^+e^- \longrightarrow \pi^+\pi^-$ cross section measurements. $\sigma_{\pi\pi}$ is the $e^+e^- \longrightarrow \pi^+\pi^-$ cross section with the radiative corrections due to the initial- and final-state radiation taken into account, δ_{rad} is the radiative correction due to the initial- and final-state radiation, σ_0 and $|F_\pi|^2$ are the cross section and the form factor of the $e^+e^- \longrightarrow \pi^+\pi^-$ process after the radiative corrections were undressed, and $\sigma_{\pi\pi}^{pol}$ is the $e^+e^- \longrightarrow \pi^+\pi^-$ undressed cross section without vacuum polarization but with the final-state radiation. Only uncorrelated errors are shown. The correlated systematic error σ_{sys} is 1.3% for $\sqrt{s} < 420$ MeV and 3.2% for $\sqrt{s} < 420$ MeV

$\sqrt{s}$, MeV	$\sigma_{\pi\pi}$, nb	δ_{rad}	σ_0, nb	$\lvert F_\pi\rvert^2$	$\sigma_{\pi\pi}^{pol}$, nb
970.0	118.12 ± 2.76	1.491	79.20 ± 1.85	3.91 ± 0.09	7.53 ± 1.81
958.0	137.16 ± 2.94	1.454	94.34 ± 2.02	4.56 ± 0.10	92.16 ± 1.97
950.0	150.02 ± 2.85	1.430	104.88 ± 1.99	4.99 ± 0.09	102.35 ± 1.94
940.0	166.55 ± 2.27	1.400	119.00 ± 1.62	5.56 ± 0.08	116.01 ± 1.58
920.0	204.99 ± 7.14	1.340	152.96 ± 5.33	6.89 ± 0.24	148.60 ± 5.18
880.0	310.82 ± 3.52	1.220	254.67 ± 2.88	10.65 ± 0.12	245.94 ± 2.78
840.0	513.80 ± 4.76	1.106	464.48 ± 4.30	17.99 ± 0.17	446.64 ± 4.13
820.0	676.03 ± 5.99	1.055	640.60 ± 5.68	23.86 ± 0.21	614.57 ± 5.45
810.0	760.19 ± 6.58	1.032	736.34 ± 6.37	26.90 ± 0.23	704.79 ± 6.10
800.0	856.66 ± 7.32	1.013	845.61 ± 7.23	30.28 ± 0.26	807.33 ± 6.90
794.0	890.86 ± 7.43	1.009	883.09 ± 7.37	31.25 ± 0.26	838.38 ± 7.00
790.0	892.35 ± 17.70	1.015	879.09 ± 17.44	30.86 ± 0.61	829.16 ± 16.45
786.0	926.47 ± 7.84	1.031	898.19 ± 7.60	31.28 ± 0.26	842.92 ± 7.13
785.0	941.34 ± 9.33	1.032	911.99 ± 9.04	31.70 ± 0.31	858.12 ± 8.51
784.0	989.76 ± 20.12	1.025	966.05 ± 19.64	33.51 ± 0.68	915.22 ± 18.61
783.0	1060.12 ± 11.38	1.010	1050.08 ± 11.27	36.35 ± 0.39	1005.99 ± 10.80
782.0	1123.55 ± 26.83	0.989	1136.34 ± 27.14	39.26 ± 0.94	1102.62 ± 26.33
781.0	1158.03 ± 10.80	0.971	1192.83 ± 11.12	41.13 ± 0.38	1169.48 ± 10.90
780.0	1211.67 ± 9.98	0.957	1266.56 ± 10.43	43.59 ± 0.36	1252.62 ± 10.32
778.0	1273.38 ± 9.47	0.944	1349.27 ± 10.03	46.25 ± 0.34	1343.80 ± 9.99
774.0	1282.06 ± 9.49	0.938	1366.85 ± 10.12	46.48 ± 0.34	1361.99 ± 10.08
770.0	1249.25 ± 9.26	0.935	1336.51 ± 9.91	45.08 ± 0.33	1330.42 ± 9.86
764.0	1247.24 ± 9.35	0.932	1338.62 ± 10.04	44.61 ± 0.33	1331.35 ± 9.99
760.0	1244.74 ± 9.58	0.927	1342.60 ± 10.33	44.39 ± 0.34	1335.30 ± 10.27
750.0	1219.07 ± 21.50	0.920	1325.56 ± 23.38	42.95 ± 0.76	1321.82 ± 23.31
720.0	989.95 ± 6.62	0.910	1087.59 ± 7.27	33.15 ± 0.22	1091.88 ± 7.30
690.0	717.99 ± 7.78	0.915	784.79 ± 8.50	22.50 ± 0.24	789.95 ± 8.56
660.0	515.95 ± 5.87	0.923	558.83 ± 6.36	15.07 ± 0.17	561.19 ± 6.39
630.0	382.69 ± 8.35	0.933	410.32 ± 8.95	10.41 ± 0.23	411.22 ± 8.97
600.0	287.18 ± 10.56	0.940	305.50 ± 11.23	7.30 ± 0.27	305.61 ± 11.23
580.0	255.24 ± 14.39	0.945	270.24 ± 15.24	6.22 ± 0.35	269.85 ± 15.22
560.0	226.60 ± 12.41	0.948	239.01 ± 13.09	5.30 ± 0.29	238.63 ± 13.07
550.0	217.52 ± 17.51	0.950	228.99 ± 18.43	4.99 ± 0.40	228.29 ± 18.37
540.0	212.67 ± 13.55	0.952	223.47 ± 14.24	4.78 ± 0.30	222.82 ± 14.20
530.0	200.04 ± 22.75	0.953	210.00 ± 23.88	4.42 ± 0.50	209.43 ± 23.82
520.0	178.13 ± 10.25	0.954	186.73 ± 10.75	3.87 ± 0.22	186.26 ± 10.72
510.0	174.28 ± 16.65	0.954	182.60 ± 17.45	3.73 ± 0.36	181.82 ± 17.38
500.0	175.22 ± 10.78	0.955	183.52 ± 11.29	3.70 ± 0.23	182.77 ± 11.24
480.0	165.18 ± 9.58	0.955	172.90 ± 10.03	3.41 ± 0.20	172.29 ± 9.99
470.0	143.94 ± 13.21	0.955	150.71 ± 13.83	2.94 ± 0.27	150.22 ± 13.78
450.0	141.32 ± 14.21	0.954	148.10 ± 14.89	2.86 ± 0.29	147.42 ± 14.82
440.0	116.15 ± 15.58	0.953	121.86 ± 16.35	2.35 ± 0.32	121.34 ± 16.28
430.0	111.27 ± 12.60	0.952	116.86 ± 13.23	2.26 ± 0.26	116.41 ± 13.18
410.0	127.38 ± 19.11	0.949	134.23 ± 20.14	2.64 ± 0.40	133.84 ± 20.08
390.0	121.81 ± 22.48	0.944	128.98 ± 23.80	2.65 ± 0.49	128.76 ± 23.76

$\sigma_{sys} = 1.3\%$ at $\sqrt{s} \geq 420$ MeV and $\sigma_{sys} = 3.2\%$ at $\sqrt{s} < 420$ MeV.

4. THE $e^+e^- \longrightarrow \pi^+\pi^-$ CROSS SECTION ANALYSIS

4.1. Theoretical Framework

In the framework of the vector meson dominance model, the cross section of the $e^+e^- \longrightarrow \pi^+\pi^-$ process is

$$\sigma_{\pi\pi}(s) = \frac{4\pi\alpha^2}{s^{3/2}} P_{\pi\pi}(s)|A_{\pi\pi}(s)|^2. \quad (19)$$

Here, $P_{\pi\pi}(s)$ is the phase space factor:

$$P_{\pi\pi}(s) = q_\pi^3(s), \quad q_\pi(s) = \frac{1}{2}\sqrt{s - 4m_\pi^2}.$$

The amplitudes of the $\gamma^* \longrightarrow \pi^+\pi^-$ transition are given by

$$|A_{\pi\pi}(s)|^2 = \left| \sqrt{\frac{3}{2}} \frac{1}{\alpha} \sum_{V = \rho, \omega, \rho', \rho''} \frac{\Gamma_V m_V^3 \sqrt{m_V \sigma(V \longrightarrow \pi^+\pi^-)}}{D_V(s)} \frac{e^{i\phi_{\rho V}}}{\sqrt{q_\pi^3(m_V)}} \right|^2, \quad (20)$$

where

$$D_V(s) = m_V^2 - s - i\sqrt{s}\Gamma_V(s),$$

$$\Gamma_V(s) = \sum_f \Gamma(V \longrightarrow f, s).$$

Here, f denotes the final state of the V vector meson decay, m_V is the vector meson mass, and $\Gamma_V = \Gamma_V(m_V)$. The following forms of the energy dependence of the vector meson total widths were used:

$$\Gamma_\omega(s) = \frac{m_\omega^2}{s}\frac{q_\pi^3(s)}{q_\pi^3(m_\omega)}\Gamma_\omega B(\omega \longrightarrow \pi^+\pi^-) + \frac{q_{\pi\gamma}^3(s)}{q_{\pi\gamma}^3(m_\omega)}\Gamma_\omega B(\omega \longrightarrow \pi^0\gamma) + \frac{W_{\rho\pi}(s)}{W_{\rho\pi}(m_\omega)}\Gamma_\omega B(\omega \longrightarrow 3\pi),$$

$$\Gamma_V(s) = \frac{m_V^2}{s}\frac{q_\pi^3(s)}{q_\pi^3(m_V)}\Gamma_V, \quad V = \rho, \rho', \rho''.$$

Here,

$$q_{\pi\gamma} = \frac{s - m_\pi^2}{2\sqrt{s}},$$

$W_{\rho\pi}(s)$ is the phase-space factor for the $\rho\pi \longrightarrow \pi^+\pi^-\pi^0$ final state [30–32]. In the energy dependence of the ρ, ρ', ρ'' mesons widths, only the $V \longrightarrow \pi^+\pi^-$ decays were taken into account. This approach is justified in the energy region $\sqrt{s} < 1000$ MeV. Nowadays, the ρ', ρ'' decays are rather poorly known, and therefore the same approximation was also used for fitting the data above 1000 MeV. The ω-meson mass and width were taken from the SND measurements: $m_\omega = 782.79$ MeV and $\Gamma_\omega = 8.68$ MeV [31].

Table 2. Various contributions to the systematic error of the determination of the $e^+e^- \longrightarrow \pi^+\pi^-$ cross section. σ_{sys} is the total systematic error, and $\sigma_{eff} = \sigma_{E>50} \oplus \sigma_{nucl} \oplus \sigma_\theta$ is the systematic inaccuracy of the detection efficiency determination

Error	Contribution at $\sqrt{s} \geq 420$ MeV, %	Contribution at $\sqrt{s} < 420$ MeV, %
$\sigma_{E>50}$	0.6	3.0
σ_{nucl}	0.2	0.2
σ_θ	0.8	0.8
σ_{eff}	1.0	3.1
σ_{sep}	0.5	0.5
σ_{IL}	0.5	0.5
σ_{rad}	0.2	0.2
σ_{sys}	1.3	3.2

The relative decay probabilities were calculated as

$$B(V \longrightarrow X) = \frac{\sigma(V \longrightarrow X)}{\sigma(V)},$$

$$\sigma(V) = \sum_X \sigma(V \longrightarrow X),$$

$$\sigma(V \longrightarrow X) = \frac{12\pi B(V \longrightarrow e^+e^-)B(V \longrightarrow X)}{m_V^2}.$$

In the analysis presented here, we used $\sigma(\omega \longrightarrow \pi^0\gamma) = 155.8$ nb and $\sigma(\omega \longrightarrow 3\pi) = 1615$ nb obtained in the SND experiments [31, 36].

The parameter $\phi_{\rho V}$ is the relative interference phase between the vector mesons V and ρ, and, hence, $\phi_{\rho\rho} = 0$. The phases $\phi_{\rho V}$ can deviate from 180° or 0°, and their values can be energy-dependent due to mixing between vector mesons. The phases $\phi_{\rho\rho'}$ and $\phi_{\rho\rho''}$ were fixed at 180° and 0°, because these values are consistent with the existing experimental data for the $e^+e^- \longrightarrow \pi^+\pi^-$ reaction.

Taking the ρ–ω mixing into account, the $\omega \longrightarrow \pi^+\pi^-$ and $\rho \longrightarrow \pi^+\pi^-$ transition amplitudes can be written as [37, 38]

$$A_{\omega\to\pi^+\pi^-} + A_{\rho\to\pi^+\pi^-} = \frac{g^{(0)}_{\gamma\rho}g^{(0)}_{\rho\pi\pi}}{D_\rho(s)}\left[1 - \frac{g^{(0)}_{\gamma\omega}}{g^{(0)}_{\gamma\rho}}\varepsilon(s)\right] + \frac{g^{(0)}_{\gamma\omega}g^{(0)}_{\rho\pi\pi}}{D_\omega(s)}\left[\varepsilon(s) + \frac{g^{(0)}_{\omega\pi\pi}}{g^{(0)}_{\rho\pi\pi}}\right], \quad (21)$$

where

$$\varepsilon(s) = \frac{-\Pi_{\rho\omega}}{D_\omega(s) - D_\rho(s)},$$

$$|g_{V\gamma}| = \left[\frac{3m_V^3\Gamma_V B(V \longrightarrow e^+e^-)}{4\pi\alpha}\right]^{1/2},$$

$$|g_{V\pi\pi}| = \left[\frac{6\pi m_V^2\Gamma_V B(V \longrightarrow \pi^+\pi^-)}{q_\pi^3(m_V)}\right]^{1/2}.$$

The superscript (0) denotes the coupling constants of the bare, unmixed state. $\Pi_{\rho\omega}$ is the polarization operator of the ρ–ω mixing:

$$\Pi_{\rho\omega}(s) = \mathrm{Re}(\Pi_{\rho\omega}(s)) + i\mathrm{Im}(\Pi_{\rho\omega}(s)), \quad (22)$$

and $\mathrm{Im}(\Pi_{\rho\omega}(s))$ can be written as

$$\mathrm{Im}(\Pi_{\rho\omega}(s)) = \sqrt{s}\left\{\frac{g^{(0)}_{\rho\pi\pi}g^{(0)}_{\omega\pi\pi}q_\pi^3(s)}{6\pi s} + \frac{g^{(0)}_{\rho\pi\gamma}g^{(0)}_{\omega\pi\gamma}q^3_{\pi\gamma}(s) + g^{(0)}_{\rho\eta\gamma}g^{(0)}_{\omega\eta\gamma}q^3_{\eta\gamma}(s)}{3}\right\}, \quad (23)$$

where

$$g_{VP\gamma} = \left[\frac{3\Gamma_V B(V \longrightarrow P\gamma)}{q^3_{P\gamma}(m_V)}\right]^{1/2}.$$

We neglected the contributions to $\mathrm{Im}(\Pi_{\rho\omega}(s))$ due to the *VP* intermediate state ($V = \omega, \rho$, $P = \pi, \eta$). The real part $\mathrm{Re}(\Pi_{\rho\omega}(s))$ can be represented as

$$\mathrm{Re}(\Pi_{\rho\omega}(s)) = \mathrm{Re}(\Pi^\gamma_{\rho\omega}(s)) + \mathrm{Re}(\Pi'_{\rho\omega}(s)), \quad (24)$$

where

$$\mathrm{Re}(\Pi^\gamma_{\rho\omega}(s)) = \frac{-4\pi g^{(0)}_{\rho\gamma}g^{(0)}_{\omega\gamma}}{s} \quad (25)$$

is the one-photon contribution to $\mathrm{Re}(\Pi_{\rho\omega}(s))$. We assume that the energy dependence of $\mathrm{Re}(\Pi'_{\rho\omega}(s))$ is negligible and can then be expressed by using the measured branching ratio

$$B(\omega \longrightarrow \pi^+\pi^-) = \frac{\Gamma_\rho(m_\omega)}{\Gamma_\omega}\left|\varepsilon(m_\omega) + \frac{g^{(0)}_{\omega\pi\pi}}{g^{(0)}_{\rho\pi\pi}}\right|^2 \quad (26)$$

as

$$\mathrm{Re}(\Pi'_{\rho\omega}) = \frac{4\pi g^{(0)}_{\rho\gamma}g^{(0)}_{\omega\gamma}}{m_\omega^2} + \frac{g^{(0)}_{\omega\pi\pi}}{g^{(0)}_{\rho\pi\pi}}(m_\omega^2 - m_\rho^2) + \left\{\frac{\Gamma_\omega B(\omega \longrightarrow \pi^+\pi^-)}{\Gamma_\rho(m_\omega)}|D_\omega(m_\omega) - D_\rho(m_\omega)|^2 - \left[\frac{g^{(0)}_{\rho\pi\gamma}g^{(0)}_{\omega\pi\gamma}q^3_{\pi\gamma}(m_\omega) + g^{(0)}_{\rho\eta\gamma}g^{(0)}_{\omega\eta\gamma}q^3_{\eta\gamma}(m_\omega)}{3} + \frac{g^{(0)}_{\omega\pi\pi}}{g^{(0)}_{\rho\pi\pi}}m_\omega\Gamma_\omega\right]^2\right\}^{1/2}. \quad (27)$$

Equation (21) can be rewritten as

$$A_{\omega\to\pi^+\pi^-} + A_{\rho\to\pi^+\pi^-} = \sqrt{\frac{3}{2}\frac{1}{\alpha}} \times \sum_{V=\omega,\rho}\frac{\Gamma_V m_V^3\sqrt{m_V\sigma(V \longrightarrow \pi^+\pi^-)}}{D_V(s)}\frac{f_{V\pi\pi}(s)}{\sqrt{q_\pi(m_V)}}, \quad (28)$$

where

$$f_{V\pi\pi}(s) = \frac{r_{V\pi\pi}(s)}{r_{V\pi\pi}(m_V)},$$

and

$$r_{\rho\pi\pi}(s) = 1 - \frac{g^{(0)}_{\gamma\omega}}{g^{(0)}_{\gamma\rho}}\varepsilon(s), \quad r_{\omega\pi\pi}(s) = \varepsilon(s) + \frac{g^{(0)}_{\omega\pi\pi}}{g^{(0)}_{\rho\pi\pi}}.$$

The theoretical value of the phase $\phi_{\rho\omega}$ can be calculated from the above expressions:

$$\phi_{\rho\omega} = \arg(f_{\omega\pi\pi}(m_\omega)) - \arg(f_{\rho\pi\pi}(m_\rho)) \approx 101^\circ.$$

The phase $\phi_{\rho\omega}$ is almost independent of energy. In this calculation, we assumed that the $\omega \longrightarrow \pi^+\pi^-$ transition proceeds only via the ρ–ω mixing, that is, $g^{(0)}_{\omega\pi\pi} = 0$. To determine the $g^{(0)}_{\rho\pi\pi}$, $g^{(0)}_{\gamma V}$, and $g^{(0)}_{VP\gamma}$ coupling constants, the corresponding measured decay widths were used.

4.2. Fit to the Experimental Data

The ρ' and ρ" parameters were determined from the fit to the $e^+e^- \longrightarrow \pi^+\pi^-$ cross section measured in the

energy region $\sqrt{s} < 2400$ MeV by OLYA and DM2 detectors [17, 39], together with the isovector part of the $e^+e^- \longrightarrow \pi^+\pi^-$ cross section calculated by assuming the CVC hypothesis from the spectral function of the $\tau^- \longrightarrow \pi^-\pi^0\nu_\tau$ decay measured by CLEO II [5],

$$\sigma_{\pi\pi}(m_i) = \frac{4(\pi\alpha)^2}{m_i}\frac{B(\tau \longrightarrow \pi\pi^0\nu_\tau)}{B(\tau \longrightarrow e\bar{\nu}_e\nu_\tau)}\frac{m_\tau^8}{12\pi|V_{ud}|^2 S_{EW}}\frac{1}{\times \dfrac{1}{m_i(m_\tau^2 - m_i^2)^2(m_\tau^2 + 2m_i^2)}\dfrac{1}{N}\dfrac{N_i}{\Delta m_i}}, \quad (29)$$

where m_i is the central value of the $\pi^-\pi^0$ pair invariant mass for the ith bin, Δm_i is the bin width, N_i is the number of entries in the ith bin, N is the total number of entries, $|V_{ud}|$ is the CKM matrix element, and $S_{EW} = 1.0194$ is the radiative correction [3, 5, 40].

The ρ' and ρ'' parameters obtained were used in fitting to the SND data (Table 3, Fig. 21). The free parameters of the fit were m_ρ, Γ_ρ, $\sigma(\rho \longrightarrow \pi^+\pi^-)$, $\sigma(\omega \longrightarrow \pi^+\pi^-)$, $\phi_{\rho\omega}$, and $\sigma(\rho' \longrightarrow \pi^+\pi^-)$. The first fit was performed with $\sigma(\rho'' \longrightarrow \pi^+\pi^-)$, ρ' and ρ'' masses, and widths fixed at the values obtained from the fit to the CLEO II and DM2 data. The second and third fits were done without the ρ'' meson. The ρ' mass and width were fixed by using the results of the fit to the CLEO II and DM2 data (the second variant in Table 3) and to the OLYA data (the third variant in Table 3). The values of the ρ and ω parameters exhibit a rather weak model dependence.

5. DISCUSSION

The comparison of the $e^+e^- \longrightarrow \pi^+\pi^-$ cross section obtained in the SND experiment with other results [8, 9, 17–19] is shown in Figs. 22–25. In the energy region $\sqrt{s} < 600$ MeV, all experimental data are in agreement (Fig. 22). Above 600 MeV, the OSPK (ORSAY–ACO) [8] and DM1 [9] points lie about 10% lower than the SND ones (Fig. 23). The SND cross section exceeds the OLYA and CMD measurements [17] by (6 ± 1)% in this energy region (Fig. 24). The systematic error of OLYA measurement is 4%, and the OLYA data agree with the SND result. The systematic uncertainty of the CMD result is 2%, and, hence, the difference between the SND and CMD results is about 2.5 of the joint systematic error. At the same time, the SND and CMD data below 600 MeV agree well (Fig. 22). The average deviation between CMD2 [18] and SND data is (1.4 ± 0.5)%; the systematic inaccuracies of these measurements are 0.6 and 1.3%, respectively. In the KLOE experiment at the φ-factory DAFΦNE, the form factor $|F_\pi(s)|^2$ was measured by using the "radiative return" method with the systematic error 0.9% [19]. In [19], the

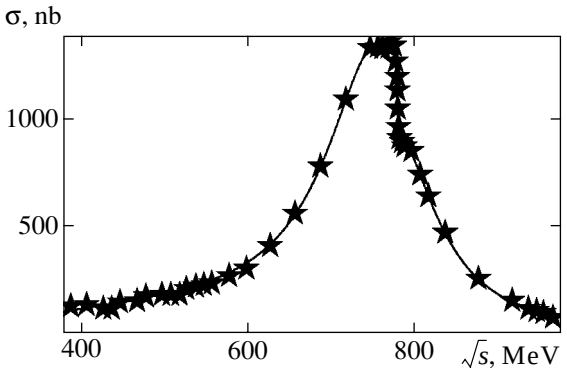


Fig. 21. The $e^+e^- \longrightarrow \pi^+\pi^-$ cross section. Stars are the SND data obtained in this work; the curve is the result of fitting.

Table 3. Fit results. The column number N corresponds to the different choices of the ρ' and ρ'' parameters

N	1	2	3
m_ρ, MeV	774.9 ± 0.4	774.9 ± 0.4	774.9 ± 0.4
Γ_ρ, MeV	146.2 ± 0.8	146.4 ± 0.8	146.3 ± 0.8
$\sigma(\rho \longrightarrow \pi^+\pi^-)$, nb	1222 ± 7	1218 ± 7	1219 ± 7
$\sigma(\omega \longrightarrow \pi^+\pi^-)$, nb	30.2 ± 1.4	30.3 ± 1.4	30.3 ± 1.4
$\varphi_{\rho\omega}$, deg	113.6 ± 1.3	113.4 ± 1.3	113.5 ± 1.3
$m_{\rho'}$, MeV	1403	1403	1360
$\Gamma_{\rho'}$, MeV	455	455	430
$\sigma(\rho' \longrightarrow \pi^+\pi^-)$, nb	3.8 ± 0.3	1.8 ± 0.2	1.9 ± 0.2
$m_{\rho''}$, MeV	1756		
$\Gamma_{\rho''}$, MeV	245		
$\sigma(\rho'' \longrightarrow \pi^+\pi^-)$, nb	1.7		
χ^2/N_{df}	50.2/39	48.8/39	49.4/39

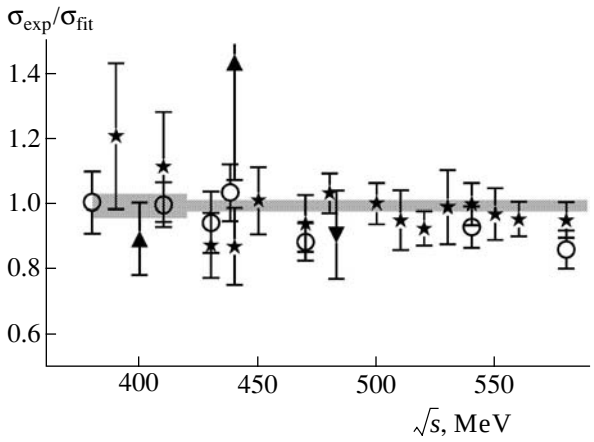


Fig. 22. The ratio of the $e^+e^- \longrightarrow \pi^+\pi^-$ cross section obtained in different experiments to the fit curve (Fig. 21). The shaded area shows the systematic error of the SND measurements. The SND (★, this work), CMD (○), OLYA (▲), and DM1 (▼) [9, 17] results are presented.

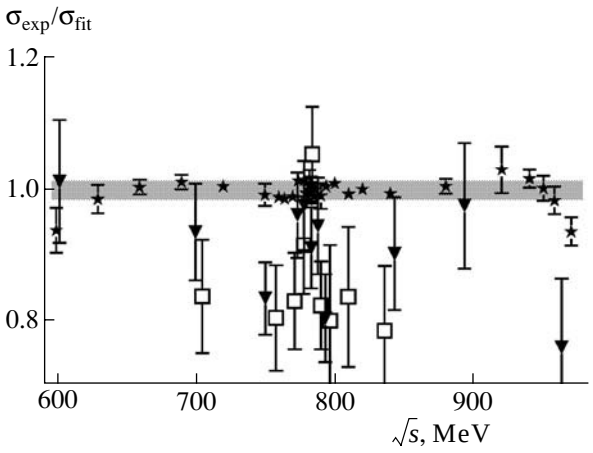


Fig. 23. The ratio of the $e^+e^- \longrightarrow \pi^+\pi^-$ cross section obtained in different experiments to the fit curve (Fig. 21). The shaded area shows the systematic error of the SND measurements. The SND (★, this work), DM1 (▼), and OSPK (□) [8, 9] results are presented.

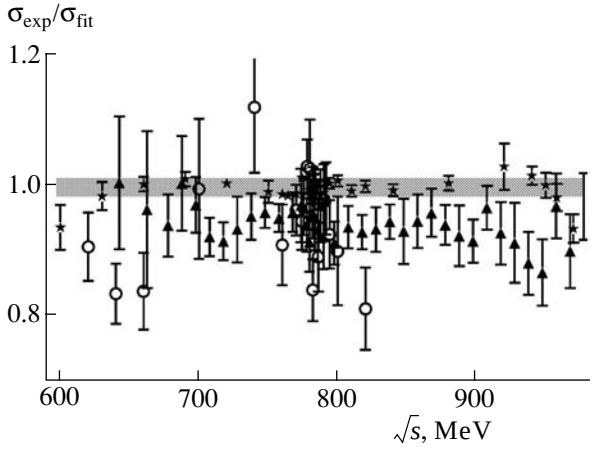


Fig. 24. The ratio of the $e^+e^- \longrightarrow \pi^+\pi^-$ cross section obtained in different experiments to the fit curve (Fig. 21). The shaded area shows the systematic error of the SND measurements. The SND (★, this work), OLYA (▲), and CMD (○) [17] results are presented.

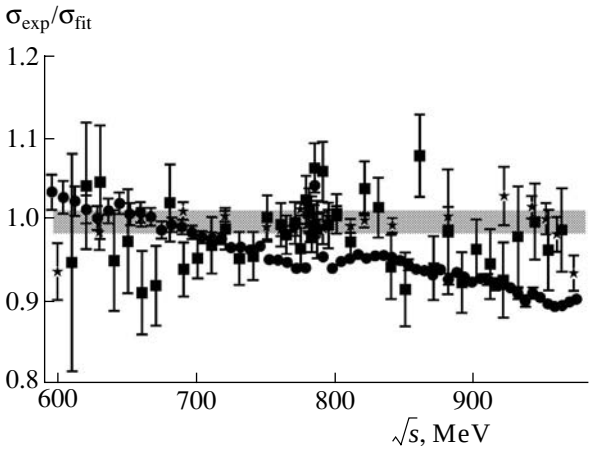


Fig. 25. The ratio of the $e^+e^- \longrightarrow \pi^+\pi^-$ cross section obtained in different experiments to the fit curve (Fig. 21). The shaded area shows the systematic error of the SND measurements. The SND (★, this work), CMD2 (■), and KLOE (●) [18, 19] results are presented.

bare form factor is listed. In order to compare the KLOE result with the SND one, we therefore appropriately dressed the form factor. The results of this comparison are shown in Fig. 25. The KLOE measurement is in conflict with the SND result as well as with the CMD2 one.

The ρ-meson parameters m_ρ, Γ_ρ, $\sigma(\rho \longrightarrow \pi^+\pi^-)$ were determined from the study of the $e^+e^- \longrightarrow \pi^+\pi^-$ cross section. The ρ meson mass and width were found to be

$$m_\rho = 774.9 \pm 0.4 \pm 0.5 \text{ MeV},$$

$$\Gamma_\rho = 146.5 \pm 0.8 \pm 1.5 \text{ MeV}.$$

The systematic error is related to the accuracy of the determination of the collider energy, to the uncertainty of the model, and to the error of determination of the cross section. The ρ-meson parameters were studied in other e^+e^- experiments by using the processes $e^+e^- \longrightarrow \pi^+\pi^-$ [17, 18], $e^+e^- \longrightarrow \rho\pi \longrightarrow \pi^+\pi^-\pi^0$ [32, 41], and the $\tau^- \longrightarrow \pi^-\pi^0\nu_\tau$ decay [3, 5]. The SND results are in agreement with these measurements, as is shown in Figs. 26 and 27.

The parameter $\sigma(\rho \longrightarrow \pi^+\pi^-)$ was found to be

$$\sigma(\rho \longrightarrow \pi^+\pi^-) = 1220 \pm 7 \pm 16 \text{ nb},$$

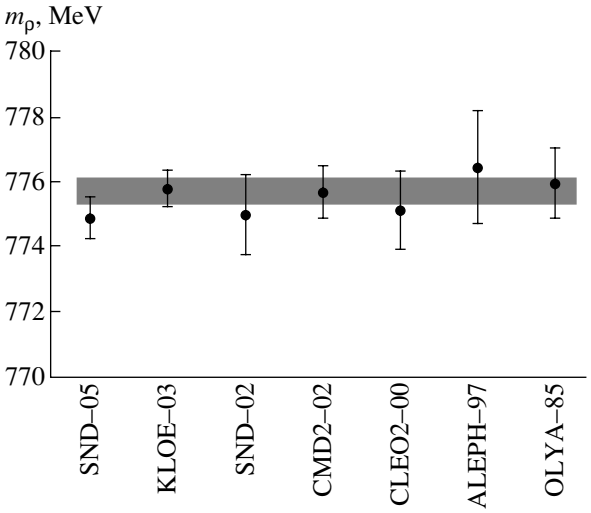


Fig. 26. The ρ-meson mass m_ρ measured in this work (SND-05) and in [3, 5, 17, 18, 32, 41]. The shaded area shows the average of the previous results.

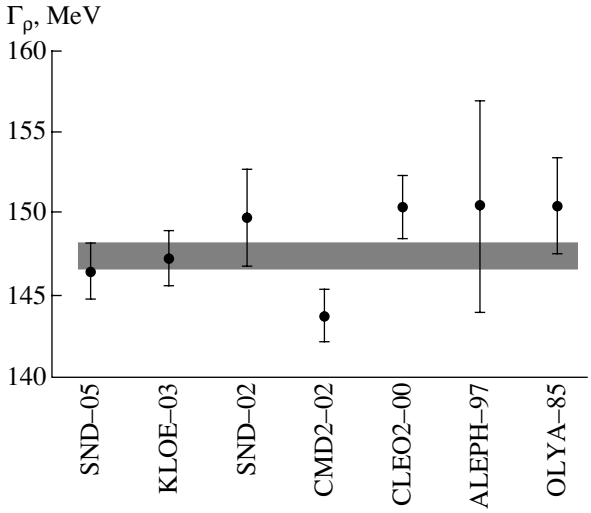


Fig. 27. The ρ meson width Γ_ρ measured in this work (SND-05) and in [3, 5, 17, 18, 32, 41]. The shaded area shows the average of the previous results.

which corresponds to

$$B(\rho \longrightarrow e^+e^-)B(\rho \longrightarrow \pi^+\pi^-) = (4.991 \pm 0.028 \pm 0.066) \times 10^{-5},$$

$$\Gamma(\rho \longrightarrow e^+e^-) = 7.31 \pm 0.021 \pm 0.11 \text{ keV}.$$

The systematic error includes systematic uncertainties in the cross section measurement and the model dependence. A comparison of the $\Gamma(\rho \longrightarrow e^+e^-)$ obtained in this work with other experimental results [8, 17, 18] and with the PDG world average [42] is shown in Fig. 28. The SND result exceeds all previous measurements. It differs by about 1.5 standard deviations from the CMD2 measurement [18] and by two standard deviations from the PDG world average [42]. The difference of the ρ-meson leptonic widths obtained by SND and CMD2 should be attributed mainly to the difference in the total widths of the ρ-meson rather than to the difference in the cross section values. The value $\sigma(\rho \longrightarrow \pi^+\pi^-) = 1198$ nb, which can be obtained by using the CMD2 cross section data reported in [18], agrees with the SND result within the measurements errors.

The parameter $\sigma(\omega \longrightarrow \pi^+\pi^-)$ was found to be

$$\sigma(\omega \longrightarrow \pi^+\pi^-) = 29.9 \pm 1.2 \pm 1.0 \text{ nb},$$

which corresponds to

$$B(\omega \longrightarrow e^+e^-)B(\omega \longrightarrow \pi^+\pi^-) = (1.247 \pm 0.062 \pm 0.042) \times 10^{-6}.$$

The systematic error is related to the model dependence, to the error of determination of the cross section, and to the accuracy of determination of the collider energy. In the previous studies of the $e^+e^- \longrightarrow \pi^+\pi^-$ reaction, the relative probability of the $\omega \longrightarrow \pi^+\pi^-$ decay was also reported. The comparison of $B(\omega \longrightarrow \pi^+\pi^-) = 0.0175 \pm 0.0011$ obtained by using the SND data and the PDG value of the $\omega \longrightarrow e^+e^-$ decay width [42] with the results of other experiments is shown in Fig. 29. The SND result is the most precise.

The phase $\phi_{\rho\omega}$ was found to be

$$\phi_{\rho\omega} = 113.5 \pm 1.3 \pm 1.7^\circ.$$

This value differs by six standard deviations from 101°, which is expected under the assumption that the $\omega \longrightarrow$

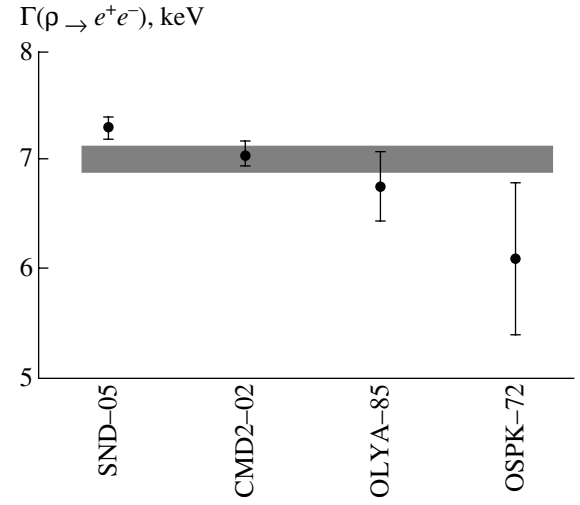


Fig. 28. The value of $\Gamma(\rho \longrightarrow e^+e^-)$ obtained in this work (SND-05) and in [8, 17, 18]. The shaded area shows the world average value [42].

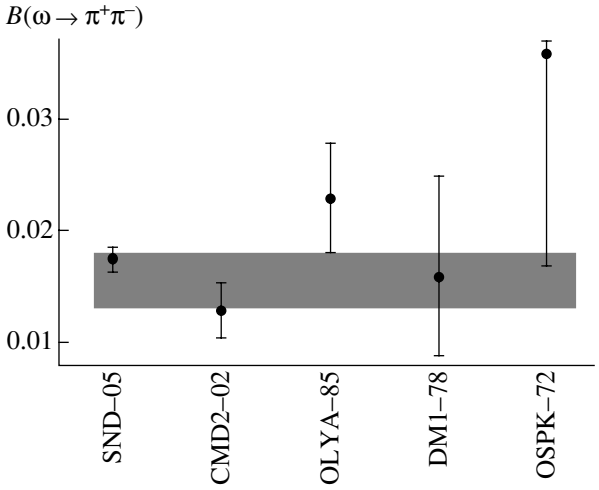


Fig. 29. The value of $B(\omega \longrightarrow \pi^+\pi^-)$ obtained in this work (SND-05) and in [8, 9, 17, 18]. The shaded area shows the world average value [42].

$\pi^+\pi^-$ transition proceeds via the ρ–ω mixing mechanism. If the ratio $g^{(0)}_{\omega\pi\pi}/g^{(0)}_{\rho\pi\pi}$ is the free parameter of the fit instead of the phase $\phi_{\rho\omega}$, it follows that

$$\frac{g^{(0)}_{\omega\pi\pi}}{g^{(0)}_{\rho\pi\pi}} = 0.11 \pm 0.01.$$

This ratio corresponds to the too large direct transition width

$$\Gamma^{(0)}(\omega \longrightarrow \pi^+\pi^-) = 1.82 \pm 0.33 \text{ MeV},$$

while the natural expectation is

$$\Gamma^{(0)}(\omega \longrightarrow \pi^+\pi^-) \approx \alpha^2\Gamma_\rho \approx 8 \text{ keV}.$$

We note that the analysis of the OLYA and CMD2 data [17, 18] gives similar values of the $\phi_{\rho\omega}$ phase. This result can point out that a considerable direct transition $\omega \longrightarrow \pi^+\pi^-$ exists. On the other hand, this discrepancy can also be attributed to inadequacies of the applied theoretical model.

The comparison of the phase $\arg(A_{\rho \to \pi^+\pi^-} + A_{\rho' \to \pi^+\pi^-})$ with the ππ scattering phase in the *P*-wave [43, 44] is shown in Fig. 30. These phases must be equal in the purely elastic scattering region. The agreement is satisfactory, and, in any case, no significant difference is observed in the energy region $\sqrt{s} \approx m_\rho$.

The comparison of the $e^+e^- \longrightarrow \pi^+\pi^-$ cross section obtained under the CVC hypothesis from the τ spectral function of the $\tau^- \longrightarrow \pi^-\pi^0\nu_\tau$ decay [3, 5] with the isovector part of the cross section measured in this work is shown in Fig. 31. The cross section obtained by SND was undressed from the vacuum polarization, and the contribution from the $\omega \longrightarrow \pi^+\pi^-$ decay was excluded. The cross section calculated from the τ spectral function was multiplied by a coefficient that takes into account the difference in the $\pi^\pm$ and π^0 masses:

$$\delta = \left(\frac{q_\pi(s)}{q_{\pi^\pm}(s)}\right)^3 \frac{\left|A_{\pi^+\pi^-}(s)\right|^2}{\left|A_{\pi^0\pi^\pm}(s)\right|^2},$$

$$q_{\pi^\pm}(s) = \frac{1}{2\sqrt{s}}[(s - (m_{\pi^0} + m_{\pi^\pm})^2) \times (s - (m_{\pi^0} - m_{\pi^\pm})^2)]^{1/2}.$$

The average deviation of the SND and τ data is about 1.5%. For almost all energy points, this deviation is

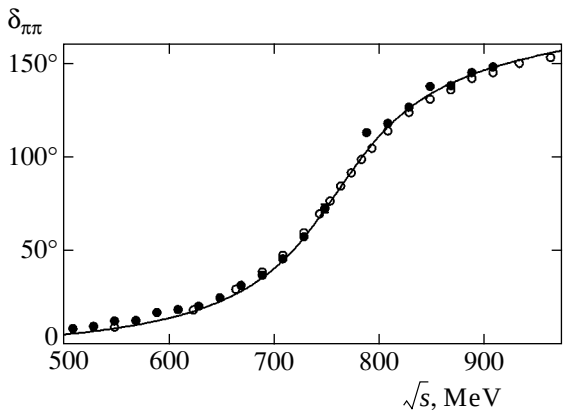


Fig. 30. The ππ scattering phase in the *P*-wave. Dots and circles are the results of the phase measurements in [43, 44] by using the reaction $\pi N \longrightarrow \pi\pi N$. The curve is the phase of the amplitude $A_{\rho \to \pi\pi} + A_{\rho \to \pi^+\pi^-}$ obtained from the fit to the SND data presented in this work.

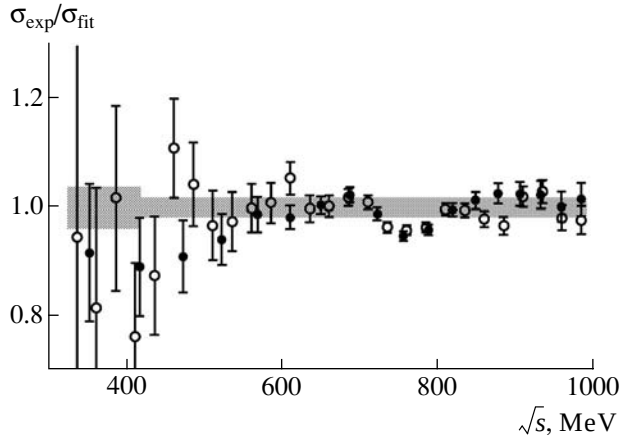


Fig. 31. The ratio of the $e^+e^- \longrightarrow \pi^+\pi^-$ cross section calculated from the $\tau^- \longrightarrow \pi^-\pi^0\nu_\tau$ decay spectral function measured in [3, 5] (○, CLEO II; ●, ALEPH) to the isovector part of the $e^+e^- \longrightarrow \pi^+\pi^-$ cross section measured in this work. The shaded area shows the joint systematic error.

within the joint systematic error (about 1.6%). The 10% difference between the e^+e^- and τ data at $\sqrt{s} > 800$ MeV, which was claimed in [45], is absent.

Using the $\sigma^{pol}_{\pi\pi}(s)$ cross section (Table 1), the contribution to the anomalous magnetic moment of the muon due to the $\pi^+\pi^-(\gamma)$ intermediate state in the vacuum polarization was calculated via the dispersion integral

$$a_\mu(\pi\pi, 390 \le \sqrt{s} \le 970 \text{ MeV}) = \left(\frac{\alpha m_\mu}{3\pi}\right)^2 \int_{s_{min}}^{s_{max}} \frac{R(s)K(s)}{s^2} ds,$$

where $s_{max} = 970$ MeV, $s_{min} = 390$ MeV, $K(s)$ is the known kernel and

$$R(s) = \frac{\sigma^{pol}_{\pi\pi}}{\sigma(e^+e^- \longrightarrow \mu^+\mu^-)},$$

$$\sigma(e^+e^- \longrightarrow \mu^+\mu^-) = \frac{4\pi\alpha^2}{3s}.$$

The integral was evaluated by using the trapezoidal rule. To take the numerical integration errors into account, the correction method suggested in [46] was applied. As a result, we obtained

$$a_\mu(\pi\pi, 390 \le \sqrt{s} \le 970 \text{ MeV}) = (488.7 \pm 2.6 \pm 6.6) \times 10^{-10}.$$

This is about 70% of the total hadronic contribution to the anomalous magnetic moment of the muon $(g-2)/2$.

If the integration is performed for the energy region corresponding to the CMD2 measurements [18], then the result is

$$a_\mu(\pi\pi) = (385.6 \pm 5.2) \times 10^{-10},$$

which is 1.8% (one standard deviation) higher than the CMD2 result:

$$a_\mu(\pi\pi) = (378.6 \pm 3.5) \times 10^{-10}.$$

Hence, no considerable difference between the SND and CMD2 results is observed.

6. CONCLUSIONS

The cross section of the process $e^+e^- \longrightarrow \pi^+\pi^-$ was measured in the SND experiment at the VEPP-2M collider in the energy region $390 < \sqrt{s} < 980$ MeV with an accuracy of 1.3% at $\sqrt{s} \ge 420$ MeV and 3.4% at $\sqrt{s} <$ 420 MeV. The measured cross section was analyzed in the framework of the generalized vector meson dominance model. The following ρ-meson parameters were obtained:

$$m_\rho = 774.9 \pm 0.4 \pm 0.5 \text{ MeV},$$

$$\Gamma_\rho = 146.5 \pm 0.8 \pm 1.5 \text{ MeV},$$

$$\sigma(\rho \longrightarrow \pi^+\pi^-) = 1220 \pm 7 \pm 16 \text{ nb}.$$

The parameters of the *G*-parity suppressed process $e^+e^- \longrightarrow \omega \longrightarrow \pi^+\pi^-$ were measured with high precision. The measured value

$$\sigma(\omega \longrightarrow \pi^+\pi^-) = 29.9 \pm 1.4 \pm 1.0 \text{ nb}$$

corresponds to the relative probability

$$B(\omega \longrightarrow \pi^+\pi^-) = 1.75 \pm 0.11\%.$$

The relative interference phase between the ρ and ω mesons was found to be

$$\phi_{\rho\omega} = 113.5 \pm 1.3 \pm 1.7^\circ.$$

This result is in conflict with the naive expectation $\phi_{\rho\omega} = 101^\circ$ from the ρ–ω mixing. The SND result agrees with the cross section calculated from the τ spectral function data within the accuracy of the measurements. Using the measured cross section, the contribution to the anomalous magnetic moment of the muon due to the $\pi^+\pi^-(\gamma)$ intermediate state in the vacuum polarization was calculated:

$$a_\mu(\pi\pi, 390 \le \sqrt{s} \le 970 \text{ MeV}) = (488.7 \pm 2.6 \pm 6.6) \times 10^{-10}.$$

ACKNOWLEDGMENTS

The authors are grateful to N.N. Achasov for useful discussions. The work is supported in part by the Russian Foundation for Basic Research, project nos. 04-02-16181a, 04-02-16184a, 05-02-16250a, and LSS-1335.2003.2.

REFERENCES

1. H. N. Brown *et al.*, Phys. Rev. Lett. **86**, 2227 (2001).
2. G. W. Bennet *et al.*, Phys. Rev. Lett. **92**, 161802 (2004).
3. R. Barate *et al.*, Z. Phys. C **76**, 15 (1997).
4. K. Ackerstaff *et al.*, Eur. Phys. J. C **7**, 571 (1999).
5. S. Anderson *et al.*, Phys. Rev. D **61**, 112002 (2000).
6. J. E. Augustin *et al.*, Phys. Rev. Lett. **20**, 126 (1968); Lett. Nuovo Cimento **2**, 214 (1969); Phys. Lett. B **28**, 508 (1969).

7. V. L. Auslender *et al.*, Phys. Lett. B **25**, 433 (1967); Yad. Fiz. **9**, 114 (1969) [Sov. J. Nucl. Phys. **9**, 69 (1969)].
8. D. Benaksas *et al.*, Phys. Lett. B **39**, 289 (1972).
9. A. Quenzer *et al.*, Phys. Lett. B **76**, 512 (1978).
10. I. B. Vasserman *et al.*, Yad. Fiz. **28**, 968 (1978) [Sov. J. Nucl. Phys. **28**, 496 (1978)].
11. A. D. Bukin *et al.*, Phys. Lett. B **73**, 226 (1978).
12. I. B. Vasserman *et al.*, Yad. Fiz. **30**, 999 (1979) [Sov. J. Nucl. Phys. **30**, 519 (1979)].
13. I. B. Vasserman *et al.*, Yad. Fiz. **33**, 709 (1981) [Sov. J. Nucl. Phys. **33**, 368 (1981)].
14. L. M. Kurdadze *et al.*, Pisma Zh. Éksp. Teor. Fiz. **37**, 613 (1983) [JETP Lett. **37**, 733 (1983)].
15. L. M. Kurdadze *et al.*, Yad. Fiz. **40**, 451 (1984) [Sov. J. Nucl. Phys. **40**, 286 (1984)].
16. S. R. Amendolia *et al.*, Phys. Lett. B **138**, 454 (1984).
17. L. M. Barkov *et al.*, Nucl. Phys. B **256**, 365 (1985).
18. R. R. Akhmetshin *et al.*, Phys. Lett. B **527**, 161 (2002); **578**, 285 (2004).
19. A. Aloisio *et al.*, Phys. Lett. B **606**, 12 (2005).
20. M. N. Achasov *et al.*, Nucl. Instrum. Methods Phys. Res. A **449**, 125 (2000).
21. A. N. Skrinsky, in *Proceedings of Workshop on Physics and Detectors for DAΦNE* (Frascati, Italy, 1995), p. 3.
22. A. D. Bukin *et al.*, in *Proceedings of Workshop on Detector and Event Simulation in High Energy Physics* (Amsterdam, 1991), p. 79.
23. A. D. Bukin and N. A. Grozina, Comput. Phys. Commun. **78**, 287 (1994).
24. K. Hänssgen and S. Ritter, Comput. Phys. Commun. **31**, 411 (1984); K. Hänssgen and J. Ranft, Comput. Phys. Commun. **39**, 37 (1986); **39**, 53 (1986); K. Hänssgen, Nucl. Sci. Eng. **95**, 137 (1987).
25. A. D. Bukin *et al.*, Report of Budker INP (Budker Inst. of Nuclear Physics, Novosibirsk, 1992), No. 92-93.
26. F. A. Berends and R. Kleiss, Nucl. Phys. B **228**, 537 (1983).
27. A. B. Arbuzov *et al.*, J. High Energy Phys. **10**, 001 (1997).
28. A. B. Arbuzov *et al.*, J. High Energy Phys. **10**, 006 (1997).
29. B. H. Denby, Comput. Phys. Commun. **49**, 429 (1988).
30. M. N. Achasov *et al.*, Phys. Rev. D **63**, 072002 (2001).
31. M. N. Achasov *et al.*, Phys. Rev. D **68**, 052006 (2003).
32. M. N. Achasov *et al.*, Phys. Rev. D **65**, 032002 (2002).
33. S. Jadach, W. Placzek, and B. F. L. Ward, Phys. Lett. B **390**, 298 (1997).
34. A. V. Bogdan *et al.*, Report of Budker INP (Budker Inst. of Nuclear Physics, Novosibirsk, 2005), No. 2005-33.
35. J. Schwinger, *Particles, Sources and Fields* (Addison-Wesley, Reading, Mass., 1973; Mir, Moscow, 1976), Vol. 2; M. Deers and K. Hikasa, Phys. Lett. B **252**, 127 (1990); K. Melnikov, Int. J. Mod. Phys. A **16**, 4591 (2001); A. Hoefer, J. Gluza, and F. Jegerlehner, Eur. Phys. J. C **24**, 51 (2002).
36. M. N. Achasov *et al.*, Phys. Lett. B **559**, 171 (2003).
37. N. N. Achasov, A. A. Kozhevnikov, and G. N. Shestakov, Phys. Lett. B **50**, 448 (1974); N. N. Achasov, N. M. Budnev, A. A. Kozhevnikov, and G. N. Shestakov, Yad. Fiz. **23**, 610 (1976) [Sov. J. Nucl. Phys. **23**, 320 (1976)]; N. N. Achasov and G. N. Shestakov, Fiz. Élem. Chastits At. Yadra **9**, 48 (1978) [Sov. J. Part. Nucl. **9**, 19 (1978)].
38. N. N. Achasov and A. A. Kozhevnikov, Yad. Fiz. **55**, 809 (1992) [Sov. J. Nucl. Phys. **55**, 449 (1992)]; Int. J. Mod. Phys. A **7**, 4825 (1992).
39. D. Bisello *et al.*, Phys. Lett. B **220**, 321 (1989).
40. E. Braaten, S. Narison, and A. Pich, Nucl. Phys. B **373**, 581 (1992).
41. A. Aloisio *et al.*, Phys. Lett. B **561**, 55 (2003).
42. S. Eidelman *et al.*, Phys. Lett. B **592**, 1 (2004).
43. P. Estabrooks and A. D. Martin, Nucl. Phys. B **79**, 301 (1974).
44. S. D. Protopopescu *et al.*, Phys. Rev. D **7**, 1279 (1973).
45. M. Davier *et al.*, Eur. Phys. J. C **27**, 497 (2003); **31**, 503 (2003).
46. N. N. Achasov and A. V. Kiselev, Pis'ma Zh. Éksp. Teor. Fiz. **75**, 643 (2002) [JETP Lett. **75**, 527 (2002)].

Journal of Experimental and Theoretical Physics, Vol. 101, No. 6, 2005, pp. 1071–1076.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 6, 2005, pp. 1220–1226.

NUCLEI, PARTICLES, FIELDS, GRAVITATION, AND ASTROPHYSICS

The Gravitational Field of a Point Charge and Finiteness of Self-Energy

M. B. Golubev[a] and S. R. Kelner[b]

[a]*Russian Federal Nuclear Center, All-Russia Research Institute of Experimental Physics, Sarov, 607195 Russia*
[b]*Moscow Institute of Engineering Physics, Kashirskoe sh. 31, Moscow, 115409 Russia*
e-mail: skelner@rambler.ru

Received July 20, 2005

Abstract—Singularities in the metric of classical solutions of the Einstein equations (the Schwarzschild, Kerr, Reissner–Nordstroem, and Kerr–Newman solutions) give rise to generalized functions in the Einstein tensor. A technique based on the limiting sequence of solutions is used to analyze these functions, which can have a more complex behavior than the Dirac δ function. We show that the solutions will satisfy the Einstein equations everywhere if the energy–momentum tensor has an appropriate singular addition of nonelectromagnetic origin. When this addition term is included, the total energy turns out to be finite and equal to mc^2, while the angular momentum for the Kerr and Kerr–Newman solutions is $mc\mathbf{a}$. Since the Reissner–Nordstroem and Kerr–Newman solutions correspond to a point charge in classical electrodynamics, the result allows us to take a fresh look at the divergence of the self-energy of a point charge. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

One of the main principles of general relativity is the equality of the inertial and gravitational masses. However, at first glance, the classical (Schwarzschild, Kerr, Reissner–Nordstroem, and Kerr–Newman) solutions of the Einstein equations do not satisfy this principle. For the Schwarzschild and Kerr solutions, the energy–momentum tensor and, hence, the self-energy are equal to zero; for the Reissner–Nordstroem and Kerr–Newman solutions, the self-energy is infinite, while the gravitational mass is finite for all these solutions.

The fact that the above solutions do not satisfy the Einstein equations in the entire space may be responsible for this inconsistency. A common property of the solutions is the presence of $1/r$ and $1/r^2$ singularities in the metric. This suggests that the Einstein tensor, which depends on the second derivatives of the metric, may contain generalized functions that are lost after direct differentiation[1] and, therefore, are disregarded in the energy–momentum tensor. Previously, this question was explored for the solutions mentioned above in the Kerr–Schild representation [1, 2]. It was shown in these papers that the Einstein tensor actually contains generalized functions that can have a more complex behavior than the Dirac δ function.

The self-energy and the angular momentum can be defined in an invariant way. We supplemented the studies [1, 2] by considering other representations of the solutions and calculated the total self-energy and the angular momentum. Singular additions to the energy–momentum tensor lead to finite energy and angular momentum. Possible physical causes of the finiteness of the self-energy in general relativity were also considered in [3].

2. ANALOGY WITH ELECTROSTATICS

A method that allows one to determine whether a generalized function appears when differentiating a singular function is easiest to explain using electrostatics as an example. The potential of a point charge

$$\varphi = \frac{e}{r} \tag{1}$$

satisfies the Poisson equation

$$\Delta\varphi = -4\pi\rho, \tag{2}$$

where $\rho = e\delta(\mathbf{r})$. One way to make sure that this is the case consists in the following. Let us substitute potential (1) with a nonsingular function of the form

$$\tilde{\varphi} = \frac{e}{r}\theta(r - r_0) + \left(\frac{3e}{2r_0} - \frac{er^2}{2r_0^3}\right)\theta(r_0 - r), \tag{3}$$

where $\theta(x)$ is the Heaviside step function ($\theta(x) = 1$ for $x > 0$ and $\theta(x) = 0$ for $x < 0$). Substituting this potential in (2), we find that $\tilde{\varphi}$ is the solution of the Poisson equation for the charge density

$$\rho = -\frac{\Delta\tilde{\varphi}}{4\pi} = \frac{3e}{4\pi r_0^3}\theta(r_0 - r). \tag{4}$$

[1] For example, in electrostatics, the potential of a point charge is $\Delta\varphi = -4\pi e\delta(\mathbf{r})$, while direct differentiation yields $\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\varphi}{\partial r}\right) = 0$.

The volume integral of (4) does not depend on r_0 and is equal to e. If we pass to the limit $r_0 \longrightarrow 0$, then

$$\tilde{\varphi} \longrightarrow e/r, \quad \rho \longrightarrow e\delta(\mathbf{r});$$

i.e., the limit of solution (3) corresponds to the presence of a point source with charge e at the coordinate origin and is the solution of Eq. (2). It is easy to show that this result does not depend on the chosen potential in the region $r < r_0$; the behavior of the potential at $r = r_0$ is not necessarily smooth. The result is always the same: in the limit $r_0 \longrightarrow 0$, the potential is $\varphi = e/r$ and the charge density is $\rho = e\delta(\mathbf{r})$.

A method of describing point sources that is simpler than the method presented above, more specifically, the Fourier transform, is commonly used in classical electrodynamics. This method is efficient in electrodynamics, because the Maxwell equations are linear. However, in the case of general relativity, the equations are nonlinear, and probably the simplest approach to studying point objects consists in the solution smoothing procedure described above. Below, we apply this procedure to the classical solutions of the Einstein equations.

3. SELF-ENERGY

What the self-energy means in general relativity is not a trivial question. In general, this question is solved using the energy–momentum pseudotensor (see, e.g., [4] and references therein). The shortcoming of this approach is that the definition of the self-energy of a system is tied to a special (Cartesian) coordinate system and is not invariant relative to coordinate transformations. The energy–momentum pseudotensor allows an energy density to be assigned to the gravitational field, which, however, cannot be localized.

The self-energy can be defined using the energy–momentum tensor only for fields and matter. For stationary or static solutions, there is a Killing vector, $\xi = \partial/\partial t$ that generates a conserved current,

$$J^i = T^i_k \xi^k, \tag{5}$$

where T^i_k is the energy–momentum tensor and $\xi^k = (1, 0, 0, 0)$ are the contravariant components of the vector ξ [5]. Since $\nabla_i J^i = 0$, the following conservation law holds:

$$\frac{d}{dt}\int d^3x\sqrt{-g}J^0 = -\int dS_\alpha \sqrt{-g}J^\alpha_0. \tag{6}$$

If the energy density is defined as the zero component of this current, then the total energy

$$E = \int d^3x\sqrt{-g}J^0 = \int d^3x\sqrt{-g}T^0_k\xi^k \tag{7}$$

will not depend on the coordinates used.

4. REISSNER–NORDSTROEM METRIC

The Reissner–Nordstroem solution [6] in Cartesian coordinates can be represented as

$$ds^2 = \Phi dt^2 - \left(\frac{1}{\Phi} - 1\right)(\mathbf{n}d\mathbf{x})^2 - d\mathbf{x}^2, \tag{8}$$

where

$$\Phi = \frac{r^2 - 2mr + Q^2}{r^2}$$

(m and Q are the mass and the charge, respectively),[2] $\mathbf{n} = \mathbf{x}/r$. This solution satisfies the Einstein equations

$$G^{ik} = 8\pi T^{ik}, \tag{9}$$

where

$$T^{ik} = \frac{1}{4\pi}\left(F^i_l F^{lk} + \frac{1}{4}g^{ik}F_{lm}F^{lm}\right)$$

is the energy–momentum tensor of the electromagnetic field everywhere, except the point $r = 0$ at which the solution is singular. The structure of the singularity of the tensor G^{ik} and the behavior of the emerging generalized function can be clarified using a procedure similar to that described in Section 2.

Let us consider a metric of form (8) with the following continuous function substituted for Φ:

$$\begin{aligned}\tilde{\Phi} &= \frac{1}{r^2}(r^2 - 2mr + Q^2)\theta(r - r_0) \\ &+ \frac{1}{r_0^2}(r_0^2 - 2mr_0 + Q^2)\theta(r_0 - r).\end{aligned} \tag{10}$$

In this case, the metric becomes nonsingular and transforms to metric (8) in the limit $r_0 \longrightarrow 0$. The requirement that the metric be continuous is necessary, since the first derivatives of the metric tensor appear in the Einstein equations nonlinearly. If we admit of discontinuities in the functions g_{ik}, then the squares of the δ functions will appear in the equations. At the same time, the second derivatives appear in the equations linearly; therefore, we may admit of discontinuities in the first derivatives if we understand the next differentiation in the sense of generalized functions.

[2] We use the units in which the gravitational constant and the speed of light are equal to 1.

The energy–momentum tensor corresponding to the smoothed metric can be derived from the Einstein equations. The (0, 0) component of this tensor is

$$T_0^0 = \frac{1}{8\pi}G_0^0 = \frac{Q^2}{8\pi r^4}\theta(r-r_0) + \left(\frac{m}{4\pi r^2 r_0} - \frac{Q^2}{8\pi r^2 r_0^2}\right)\theta(r_0-r). \quad (11)$$

The first term on the right-hand side is the energy density of the electrostatic field in the region $r > r_0$. The second term, which arose from the metric smoothing, does not vanish in the limit $r_0 \longrightarrow 0$. The self-energy of the constructed solution is

$$E = \frac{Q^2}{2r_0} + \left(m - \frac{Q^2}{2r_0}\right) = m. \quad (12)$$

It can be shown that result (12) does not depend on the metric smoothing method. In the limit $r_0 \longrightarrow 0$, Eq. (11) can be written as

$$T_0^0 = m\delta(\mathbf{r}) + \frac{1}{2}Q^2\varpi(\mathbf{r}). \quad (13)$$

Here, $\varpi(\mathbf{r})$ is a generalized function defined by the following integration rule:

$$\int f(\mathbf{r})\varpi(\mathbf{r})d^3x = \int\frac{f(\mathbf{r}) - f(0)}{4\pi r^4}d^3x, \quad (14)$$

where $f(\mathbf{r})$ is a bounded smooth function. The limiting expression (13) does not depend on the metric smoothing method either. The Fourier transform of the function $\varpi(\mathbf{r})$ is

$$\tilde{\varpi}(\mathbf{k}) \equiv \int\varpi(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}}d^3x = -\frac{\pi}{4}|\mathbf{k}|. \quad (15)$$

The function

$$\varpi(\mathbf{r}) = \frac{1}{4\pi r^4}$$

everywhere, except the point $\mathbf{x} = 0$; therefore, the quantity $(1/2)Q^2\omega(\mathbf{r})$ in Eq. (13) describes the energy density of the electrostatic field. Since there is a singularity at the coordinate origin, $\tilde{\varpi}(\mathbf{k} = 0) = 0$; i.e., the integral of $\varpi(\mathbf{r})$ is zero throughout the space. This implies that the divergence of the electrostatic field energy is offset by an infinite negative energy at the center.

The remaining components of the energy–momentum tensor in the limit $r_0 \longrightarrow 0$ are

$$T_\alpha^0 = 0, \quad (16)$$

$$T_\beta^\alpha = \frac{1}{2}[m(3n_\alpha n_\beta - \delta_{\alpha\beta})\delta(\mathbf{r}) + Q^2(2n_\alpha n_\beta - \delta_{\alpha\beta})\varpi(\mathbf{r})]. \quad (17)$$

The Fourier transform of the spatial energy–momentum tensor components is

$$\tilde{T}_\beta^\alpha = \frac{\pi Q^2 k}{16}\left(\delta_{\alpha\beta} - \frac{k_\alpha k_\beta}{k^2}\right). \quad (18)$$

For the Schwarzschild metric ($Q = 0$ in (8)), the term $m\delta(\mathbf{r})$ in T_k^i, which corresponds to a point charge, can be obtained directly if the presence of the term $\Delta(1/r)$ in G_k^i is taken into account. A more complex generalized function $\varpi(\mathbf{r})$ emerges as the source in the case of $Q \neq 0$. It owes its origin to the fact that $\Delta(1/r^2)$ appears in G_k^i.

Thus, the Schwarzschild and Reissner–Nordstroem solutions can be extended to the entire space if a point source is added to the energy–momentum tensor. This assertion can be applied both to a point charge and to a black hole, since the charge–mass relation was used nowhere.

In standard units, Eq. (13) can be written as

$$T_0^0 = mc^2\delta(\mathbf{r}) + \frac{1}{2}Q^2\varpi(\mathbf{r}), \quad (19)$$

where $\varpi(\mathbf{r})$ is defined in (14). Note that the gravitational constant does not appear in this expression. It is easy to verify that the limiting formulas (16)–(18) for the energy–momentum tensor do not depend on the gravitational constant either. Therefore, the results are also valid for flat space–time. Thus, general relativity allows the energy–momentum tensor of a stationary (or uniformly moving) point charge to be found in classical electrodynamics. In this case, the total energy of the charge is finite.

5. KERR–NEWMAN METRIC

Let us use the same procedure to analyze the structure of the singularity in the Kerr–Newman metric [7]:

$$ds^2 = \eta_{ik}dx^i dx^k + \Psi\left(dt - \frac{(r^2x_\alpha + r[\mathbf{x}\times\mathbf{a}]_\alpha + a_\alpha(\mathbf{a}\cdot\mathbf{x}))dx^\alpha}{r(r^2+a^2)}\right)^2. \quad (20)$$

Here,

$$\Psi = \frac{Q^2 - 2mr}{r^2 + \dfrac{(\mathbf{a}\cdot\mathbf{x})^2}{r^2}}, \quad (21)$$

where $\mathbf{a}$ is a space vector, a is its magnitude, and r is

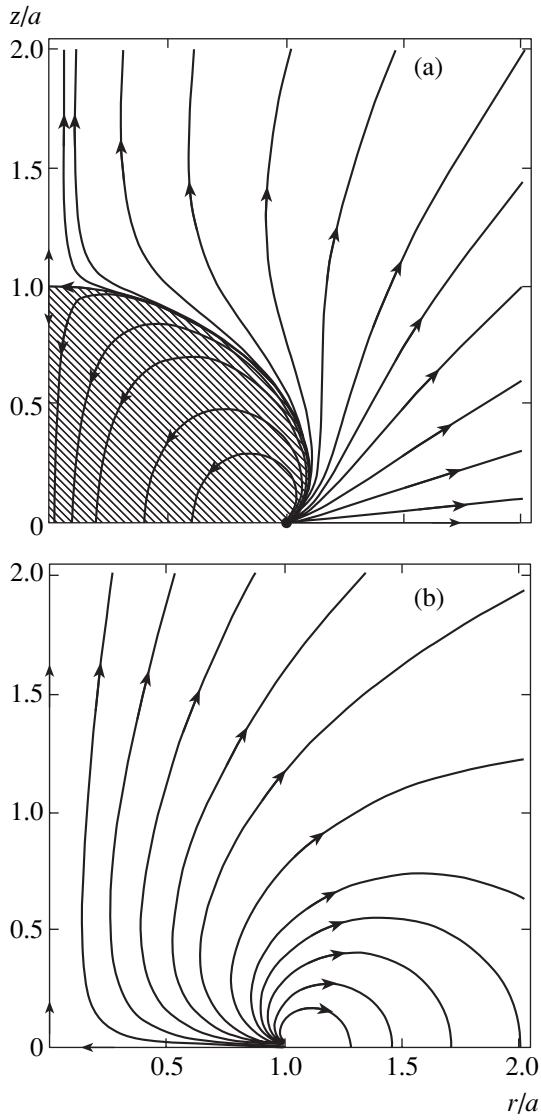


Fig. 1. Electric (a) and magnetic (b) field lines for the Kerr-Newman solution. The region of closed electric field lines is hatched.

defined by the equation

$$r^4 - r^2(\mathbf{x}^2 - a^2) - (\mathbf{a}\cdot\mathbf{x})^2 = 0. \tag{22}$$

It is easy to show that this is the standard Kerr–Schild representation if the vector **a** is directed along the *z* axis. The surfaces of constant *r* are ellipsoids of revolution whose axis coincides with the direction of the vector **a**. At $r = 0$, the ellipsoid degenerates into a disk of radius *a*. On this disk, the metric is continuous, but the components of the metric and the 4-potential of the field have a kink, while the electromagnetic field strengths have a discontinuity (see the figure). This implies that there is a singular distribution of mass, charge, and currents on the disk that is not embodied in the energy–momentum tensor of the electromagnetic field.

Let us construct a solution similar to solution (10) by substituting Ψ with the function

$$\tilde{\Psi} = \frac{Q^2 - 2mr}{r^2 + \dfrac{(\mathbf{a}\cdot\mathbf{x})^2}{r^2}}\theta(r - r_0) + \frac{Q^2 - 2mr_0}{r_0^2 + \dfrac{(\mathbf{a}\cdot\mathbf{x})^2}{r_0^2}}\theta(r_0 - r). \tag{23}$$

The constructed solution is continuous everywhere, but its derivative has a discontinuity at $r = r_0$. To satisfy the Einstein equations, the (0, 0) component of the energy–momentum tensor must have the form

$$\begin{aligned} T_0^0 = \frac{1}{8\pi}G_0^0 = \frac{Q^2}{4\pi}\left(\frac{a^2}{\rho^3} - \frac{r^2}{\rho^3} - \frac{1}{2\rho^2}\right)\theta(r - r_0) \\ + \frac{r_0^2(2Mr_0 - Q^2)(r_0^4 - 3(\mathbf{a}\cdot\mathbf{x})^2)}{8\pi r^2(r^4 + (\mathbf{a}\cdot\mathbf{x})^2)} \\ \times \frac{(r^4 r_0^4 - (\mathbf{a}\cdot\mathbf{x})^4 + r^6 a^2 + r^2(\mathbf{a}\cdot\mathbf{x})^2 a^2)}{(r_0^4 + (\mathbf{a}\cdot\mathbf{x})^2)^3}\theta(r_0 - r) \\ + \frac{r_0(Mr_0(r_0^4 - 3(\mathbf{a}\cdot\mathbf{x})^2) - Q^2(r_0^4 - (\mathbf{a}\cdot\mathbf{x})^2))}{8\pi(r_0^4 + (\mathbf{a}\cdot\mathbf{x})^2)^3} \\ \times (r_0^2 a^2 - (\mathbf{a}\cdot\mathbf{x})^2)\delta(r - r_0), \end{aligned} \tag{24}$$

where

$$\rho = r^2 + (\mathbf{a}\cdot\mathbf{x})^2/r^2.$$

The contribution from each of the three parts of T_0^0 to the self-energy for $r_0 \longrightarrow 0$ is a diverging quantity, but the divergences remarkably cancel out in the sum:

$$\begin{aligned} E = \frac{Q^2}{4r_0} + \frac{Q^2(r_0^2 + a^2)\lambda}{4r_0^2 a} \\ + (2mr_0 - Q^2)\left(\frac{(5r_0^2 + a^2)\lambda}{4r_0^2 a} - \frac{3}{4r_0}\right) \\ - \frac{2Q^2 - 5mr_0}{2r_0} + \frac{(2Q^2 r_0 - m(5r_0^2 + a^2))\lambda}{2r_0 a} = m, \end{aligned} \tag{25}$$

where

$$\lambda = \arctan\left(\frac{a}{r_0}\right).$$

The first, second, and third rows of this formula are the contributions from the regions $r > r_0$, $r < r_0$, and the surface $r = r_0$, respectively. The contribution from each of the regions depends on the metric smoothing method. For example, if the functions g_{ik} are chosen in such a way that the derivatives $\partial g_{ik}/\partial r$ are continuous at r_0, then the term with the δ function will disappear in Eq. (24). However, the total energy E, as in the case of the Reissner–Nordstroem metric, does not depend on the metric smoothing method (see Appendix). The result obtained can also be extended to the Kerr metric (it will suffice to set $Q = 0$ in Eqs. (20)–(25)).

The energy–momentum tensor also allows the total angular momentum of the system to be determined. Because of the azimuthal symmetry, there is a Killing vector, $\eta = \partial/\partial\varphi$, that allows the following conserved current to be introduced:

$$I^i = -T^i_k\eta^k, \quad \nabla_i I^i = 0. \tag{26}$$

The total angular momentum is defined by the equality

$$M = \int d^3x\sqrt{-g}I^0 = \int d^3x\sqrt{-g}T^0_k\eta^k. \tag{27}$$

The contributions from the regions $r > r_0$, $r < r_0$, and the surface $r = r_0$ to integral (27) in the limit $r_0 \longrightarrow 0$ diverge, but these divergences cancel out, and the total angular momentum proves to be finite, irrespective of the metric smoothing method:

$$M = \frac{Q^2}{4}\left(\lambda\left(\frac{a}{r_0}+\frac{r_0}{a}\right)^2 - \frac{r_0}{a}+\frac{a}{r_0}\right)$$

$$+ (2mr_0 - Q^2)\left(\lambda\left(\frac{3}{2}+\frac{5r_0^2}{4a^2}+\frac{a^2}{4r_0^2}\right) - \frac{5r_0}{4a}-\frac{3a}{4r_0}\right) \tag{28}$$

$$-(a^2+r_0^2)\left(\lambda\left(\frac{m}{2r_0}+\frac{5mr_0-2Q^2}{2a^2}\right)+\frac{Q^2}{ar_0}-\frac{5m}{2a}\right) = ma.$$

In our calculations, we used the Mathematica 5 code (Wolfram Research, Inc.).

6. CONCLUSIONS

As we showed here, for the Schwarzschild, Kerr, Reissner–Nordstroem, and Kerr–Newman solutions to satisfy the Einstein equations in the entire space, including $r_0 = 0$, singular terms containing generalized functions must be added to the energy–momentum tensor. In this case, the total energy for all solutions is finite and equal to mc^2. For solutions with a nonzero charge, this addition plays the role of Poincaré tensions; i.e., an infinite negative mass that compensates for the electrostatic energy of the charge is located at the center. The emergence of a negative mass when regularizing the metric was also pointed out in [8].

The fact that the gravitational attraction for test particles transforms to repulsion even at the classical radius Q^2/mc^2 also suggests the existence of a negative mass for solutions with a nonzero charge. This can be verified by analyzing the equations of motion for test particles.

It follows from the results obtained that a self-consistent classical electron model cannot be constructed without invoking additional fields (or matter) with unusual properties. These fields must give a negative contribution to the energy.

ACKNOWLEDGMENTS

This work was supported by the International Scientific and Technical Center (ISTC) (grant no. 1655).

APPENDIX

The smoothed solutions and the corresponding total energy were analyzed in a number of papers (see, e.g., [8] and references therein). Here, we show that our definition of E (Eq. (7)) yields a result that is not only invariant relative to coordinate transformations, but also does not depend on the metric smoothing method in the important special case where the Kerr metric structure (20) is retained after smoothing.

The commutator of the covariant derivatives (see, e.g., [9]) is

$$(\nabla_k\nabla_l - \nabla_l\nabla_k)\xi_i = \xi^m R_{mikl}, \tag{29}$$

where R_{mikl} is the Riemann tensor. Multiplying this expression by g^{ik} yields the Ricci tensor on the right-hand side:

$$\nabla_i\nabla_l\xi^i - \nabla_l\nabla_i\xi^i = -\xi^m R_{ml}. \tag{30}$$

In this formula, ξ^i is an arbitrary vector. Let ξ^i be the Killing vector. The second term in (30) then vanishes, and the following relation is obtained:

$$\nabla_i\xi^{i;k} = -R^{kl}\xi_l. \tag{31}$$

Since $\xi^{i;k}$ is an antisymmetric vector, we have

$$\nabla_i\nabla_k\xi^{i;k} = \frac{1}{\sqrt{-g}}\frac{\partial^2}{\partial x^i\partial x^k}\xi^{i;k} = 0 \tag{32}$$

and arrive at the equality

$$\nabla_k(R^{kl}\xi_l) = 0. \tag{33}$$

It turns out that not only $T^{kl}\xi_l$, but also the vector $R^{kl}\xi_l$ is conserved. It thus follows that the equality

$$\nabla_k(R\xi^k) = \xi^k\frac{\partial R}{\partial x^k} = 0, \tag{34}$$

where R is the scalar curvature, is valid. However, the latter expression is obvious, since the Killing vector is a generator of transformations that do not change the metric.

It follows from equality (31) written in the form

$$R^{kl}\xi_l = \frac{1}{\sqrt{-g}}\frac{\partial}{\partial x^i}(\sqrt{-g}\xi^{i;k}) \tag{35}$$

that the quantity

$$\int d^3x\sqrt{-g}R^{kl}\xi_l = \oint dS_i\sqrt{-g}\xi^{i;k}; \tag{36}$$

thus, it does not depend on the metric smoothing method in the region $r < r_0$.

For a metric of form (20) with an arbitrary function $\Psi(r, \theta)$, the scalar curvature is

$$R = -\frac{\sin\theta}{\sqrt{-g}}\frac{\partial^2}{\partial r^2}[(r^2 + a^2\cos^2\theta)\Psi(r, \theta)]. \tag{37}$$

The formula can be verified by direct, but cumbersome calculations. Therefore, the integral of the scalar curvature appearing in the total energy,

$$-\frac{1}{2}\int R\sqrt{-g}drd\theta d\varphi = \frac{1}{2}$$
$$\times\int\frac{\partial^2}{\partial r^2}[(r^2 + a^2\cos^2\theta)\Psi(r, \theta)]dr\sin\theta d\theta d\varphi, \tag{38}$$

does not depend on the behavior of $\Psi(r, \theta)$ in the region $r < r_0$. In other words, quantity (38) does not depend on the metric smoothing method in the region $r < r_0$ either.

We only need to require that the following equality hold:

$$\left.\frac{\partial\Psi(r, \theta)}{\partial r}\right|_{r = 0} = 0.$$

It follows from Eqs. (36) and (38) that the scalar

$$\int\left(R^0_k - \frac{1}{2}g^0_k R\right)\xi^k\sqrt{-g}d^3x, \tag{39}$$

which is proportional to self-energy (7), is insensitive to the metric smoothing method.

In a similar way, we can prove that the total angular momentum (27) does not depend on the metric smoothing method either. The proof is easier, since the term with the scalar curvature does not contribute to the total angular momentum.

REFERENCES

1. H. Balasin and H. Nachbagauer, Class. Quantum Grav. **10**, 2271 (1993); gr-qc/9305009.
2. H. Balasin and H. Nachbagauer, Class. Quantum Grav. **11**, 1453 (1994); gr-qc/9312028.
3. R. Arnowitt, S. Deser, and C. Misner, in *Gravitation: An Introduction to Current Research*, Ed. by L. Witten (Wiley, New York, 1962); gr-qc/0405109.
4. S. S. Xulu, Astrophys. Space Sci. **283**, 23 (2003).
5. B. S. DeWitt, *Dynamical Theory of Groups and Fields* (Gordon and Breach, New York, 1965; Nauka, Moscow, 1987).
6. C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, New York, 1973; Mir, Moscow, 1977).
7. G. C. Debney, R. P. Kerr, and A. Schild, J. Math. Phys. **10**, 1842 (1969).
8. A. Burinskii, E. Elizalde, S. R. Hildebrandt, and G. Magli, Phys. Rev. D **65**, 064039 (2002).
9. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 7th ed. (Nauka, Moscow, 1988; Pergamon, Oxford, 1975).

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 101, No. 6, 2005, pp. 1077–1090.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 6, 2005, pp. 1227–1242.

ORDER, DISORDER, AND PHASE TRANSITIONS IN CONDENSED SYSTEMS

Magnetic Susceptibility of Quasi-One-Dimensional Ising Superantiferromagnets FeTAC and $MCl_2 \cdot 2NC_5H_5$ (M = Co, Fe): Approximation with $L \times \infty$ and $L \times L \times \infty$ Chain Clusters

M. A. Yurishchev

Institute of Problems of Chemical Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia

e-mail: yur@itp.ac.ru

Received January 18, 2005

Abstract—The temperature dependence of the zero-field susceptibilities of 2D and 3D Ising lattices with anisotropic coupling is analyzed. Infinite 2D and 3D lattices are approximated, respectively, by ensembles of independent $L \times \infty$ and $L \times L \times \infty$ chain clusters that are infinitely long in the strong-coupling (J) direction. This approach is used as a basis for a quantitative description of available experimental data on the magnetic susceptibilities of the 2D anisotropic Ising ferromagnet $[(CH_3)_3NH]FeCl_3 \cdot 2H_2O$ (FeTAC) and the quasi-one-dimensional 3D systems $CoCl_2 \cdot 2NC_5H_5$ and $FeCl_2 \cdot 2NC_5H_5$ in the entire experimental temperature range. A method is proposed for determining the relative interchain coupling strength J'/J from the maximum susceptibility value, which improves the accuracy of estimates for J'/J by more than an order of magnitude. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Various currently known materials are Ising magnets [1, 2]. Even though the Ising coupling has an extremely simple form, the macroscopic properties of these materials, such as magnetic susceptibility, are very difficult to calculate. It should be noted that no rigorous closed-form expression for the longitudinal component of susceptibility has been obtained to this day even in exactly solvable Ising models. Moreover, there are reasons to believe that no such expression can be found in the class of differentially finite (holonomic) functions [3–5] (see also [6]).

The subject of this study is the magnetic susceptibility of quasi-one-dimensional Ising magnets. Systems of two types are considered: anisotropic square lattices with coupling constants J and J' such that $|J'/J| \le 1$, and simple cubic lattices with dominant interaction along one axis represented by J and equal constants J' of interaction along the remaining two orthogonal axes.

Crystals of $[(CH_3)_3NH]FeCl_3 \cdot 2H_2O$ (FeTAC) have a 2D magnetic lattice consisting of bonded spin chains lying in a plane [7–10]. In crystals of $CoCl_2 \cdot 2NC_5H_5$ [11–14] and $FeCl_2 \cdot 2NC_5H_5$ [15–17], chains of magnetic ions are bonded into 3D systems. All of these materials are typical quasi-1D Ising superantiferromagnets that can be modeled by effective spin-1/2 Hamiltonians (with $J > 0$ and $J' < 0$). As temperature decreases, ferromagnetically ordered spin chains become antiferromagnetically ordered. Their magnetic susceptibilities have distinct maxima at temperatures T_{max} above the respective critical points T_c. The phase transition manifests itself in the susceptibility curve as an inflection point where the tangent line to the curve is infinitely steep (in the ideal case).

The susceptibility of a 2D Ising lattice was calculated in [18] for the entire temperature range (in theory, from zero to infinity). The approximation used in that study (decoupling of many-spin correlation functions) is accurate within 0.35% in the isotropic model. However, the analysis presented below shows that the error in the coordinates of the susceptibility maximum amounts to tens of percent even for $J'/J = -0.1$ ($J > 0$). Therefore, this approximation cannot be applied to quasi-1D systems in practical calculations.

The results obtained for 3D systems are even less accurate. The most reliable calculations of susceptibility for such systems make use of power series expansions. For the zero-field longitudinal susceptibility of the isotropic simple cubic Ising lattice, high-temperature expansions to the 25th- and even 32th-order terms were obtained in [19, 20] and [21], respectively. However, analogous expansions for anisotropic lattices are known only to the 10th- or 11th-order terms (see [22] and [23], respectively). Moreover, partial sums of the series rapidly diverge with increasing lattice anisotropy. In what follows, it is demonstrated that the available high-temperature series expansions of superantiferromagnetic susceptibility [24] result in unacceptably large errors for $|J'|/J = 10^{-2}$ (even after their conver-

gence is improved by Padé–Borel resummation). Note that interpretation of the experimental data discussed here requires modeling with an even smaller value of this parameter.

In this paper, susceptibilities are calculated by using cluster simulation. It is well known [25–27] that various characteristics calculated by this method (including susceptibility) converge to their values for an infinite system at an exponential rate with increasing cluster size everywhere in the parameter space except for a narrow critical region. However, this region cannot be resolved by modern experimental methods for the quasi-1D materials discussed here.

In view of the specific anisotropy to be modeled, chain clusters of infinite length in the direction of the dominant interaction J are used as subsystems ($L \times \infty$ strips and $L \times L \times \infty$ parallelepipeds for 2D and 3D systems, respectively). Undesirable surface effects are eliminated by setting periodic boundary conditions in the transverse directions for subsystems of both types. Furthermore, frustration is obviated by using chains of length $L = 2, 4, \ldots$ (measured in units of the lattice constant), with the only exception of a single chain ($L = 1$). Thus, the magnetic lattice of an $L^{d-1} \times \infty$ superantiferromagnetic cluster (in space of dimension $d = 2$ or 3) consists of two identical interpenetrating sublattices with opposite magnetic moments.

In Section 2, formulas for susceptibilities are presented, including both general expressions well suited for computations and exact asymptotic ones. The cumbersome analytical formulas derived for few-chain subsystems are relegated to the Appendix. In Section 3, the strip width ensuring the accuracy required to calculate the susceptibility of FeTAC is determined. In Section 4, the corresponding calculated results are presented. Sections 5 and 6 contain results for 3D systems analogous to those presented in the preceding two sections. Section 7 summarizes the principal results of this study.

2. CALCULATION OF SUSCEPTIBILITIES

The anisotropic Ising Hamiltonian is written as

$$H = -\frac{1}{2}J\sum_{\langle i,j\rangle}\sigma_i^z\sigma_j^z - \frac{1}{2}J'\sum_{[i,j]}\sigma_i^z\sigma_j^z, \quad (1)$$

where the Pauli matrices σ_i^z are localized at the sites of a square or simple cubic lattice. The sums with $\langle i,j\rangle$ and $[i,j]$ are taken over the nearest-neighbor pairs along the directions corresponding to J and J', respectively.

According to Kubo's linear response theory [28, 29], the static zero-field susceptibility tensor is

$$\chi_{\mu\nu} = -\beta\langle M_\nu\rangle\langle M_\mu\rangle + \int_0^\beta d\beta'\langle M_\nu(\beta')M_\mu\rangle. \quad (2)$$

Here, μ and ν stand for x, y, or z; $\beta = 1/k_BT$ is the inverse temperature measured in energy units (k_B is Boltzmann's constant); angle brackets denote ensemble-averaged quantities; M_μ is the projection of the magnetic moment of the system on the μ axis; and $M_\nu(\beta) = e^{\beta H}M_\nu e^{-\beta H}$ is a component of magnetization in the Matsubara representation.

Note that superantiferromagnets, being characterized by zero total spontaneous sublattice magnetizations, have zero magnetic moments in the absence of applied field: $\langle M_\nu\rangle = 0$. Therefore, the first term on the right-hand side of (2) vanishes under the conditions considered in this study.

The component of the magnetic moment parallel to the z axis is

$$M_z = \frac{1}{2}g_\parallel\mu_B\sum_{i=1}^{N}\sigma_i^z, \quad (3)$$

where $g_\parallel$ is the longitudinal g factor, μ_B is the Bohr magneton, and N is the total number of particles in the system. Since M_z commutes with Hamiltonian (1) and $\langle M_z\rangle = 0$, the expression for the molar zero-field longitudinal (parallel) susceptibility obtained by substituting (3) into (2) is

$$\chi_\parallel(T) \equiv \lim_{N\to\infty}\frac{N_A}{N}\chi_{zz} = \frac{N_A g_\parallel^2\mu_B^2}{4k_BT}\lim_{N\to\infty}\sum_{j=1}^{N}\langle\sigma_{i_0}^z\sigma_j^z\rangle_N, \quad (4)$$

where N_A is Avogadro's number and i_0 is any particular site in a uniform lattice ($\chi_\parallel$ is independent of its location). To evaluate the longitudinal susceptibility, one must calculate and add up all spin–spin correlation functions and take the infinite-lattice limit.

The longitudinal susceptibilities of single-, double-, and four-chain Ising models are known in analytical form (see Appendix). An analysis of these formulas shows that the predicted variation of the susceptibilities of superantiferromagnetic clusters with temperature is in qualitative agreement with experimental data. The susceptibility curve has a peak (see Fig. 1), and its magnitude indefinitely increases with lattice anisotropy. At temperatures below the maximum point, the susceptibility curve has an inflection point that approximately corresponds to the critical point of the entire system. The slope of the tangent line at the inflection point increases with the number of chains in a subsystem, approaching infinity.

For subsystems consisting of a larger number of chains, the susceptibility can be found only by numerical methods. One formula well suited for computing

the longitudinal susceptibility of an $L^{d-1} \times \infty$ Ising cluster by the transfer matrix method is [30, 31]

$$\chi_{\parallel}^{(L^{d-1}\times\infty)}(T) = \frac{N_A g_{\parallel}^2 \mu_B^2}{4L^{d-1}k_B T} \times \sum_s{}' \frac{\lambda_1 + \lambda_s}{\lambda_1 - \lambda_s} \left| f_s^+ (\hat{\sigma}_1 + \ldots + \hat{\sigma}_n) f_1 \right|^2. \quad (5)$$

Here, the primed sum skips the term with $s = 1$; $n = L^{d-1}$ is the number of chains in a cluster; λ_s and f_s denote eigenvalues and the corresponding eigenvectors of the transfer matrix, respectively (the largest eigenvalue λ_1 is nondegenerate by the Perron–Frobenius theorem [32]); and the matrix $\hat{\sigma}_k$ is defined as

$$\hat{\sigma}_k = 1 \times \ldots \times 1 \times \sigma_z \times 1 \times \ldots \times 1. \quad (6)$$

In this direct product of n matrices, the kth multiplicand is the Pauli matrix σ_z, and the remaining ones are two-dimensional identity matrices. The 2×2^n-by-2×2 transfer matrix V has the elements

$$\langle \sigma_1, \ldots, \sigma_n | V | \sigma_1', \ldots, \sigma_n' \rangle = \exp\left[K \sum_{i=1}^{n} \sigma_i \sigma_i' + \frac{1}{2} K' \sum_{[i,j]} (\sigma_i \sigma_j + \sigma_i' \sigma_j') \right], \quad (7)$$

where $\sigma_i = \pm 1$ are collinear spins in the cross section of an $L^{d-1} \times \infty$ lattice, $K = J/2k_BT$, and $K' = J'/2k_BT$. The transfer matrix V is a positive real symmetric one.

Expression (5) follows from (4); i.e., it can be derived from Kubo's linear response theory. A physically equivalent expression that has a somewhat different form was obtained by developing a perturbation series in external field for the transfer matrix [33, 34].

The key problem in evaluating the susceptibility is thus reduced to the eigenvalue–eigenvector problem for the transfer matrix V. Here, this problem is solved either by direct numerical diagonalization of the matrix V or by diagonalizing the subblocks constituting the transfer matrix block diagonalized by using cluster symmetries.

Starting again from Kubo's formula (2), one can readily show that the expression for the transverse susceptibility $\chi_\perp$ is a linear combination of a finite number of local σ^z correlation functions. These correlation functions and $\chi_\perp$ have been calculated only for isotropic 2D Ising lattices [35–37].

The available analytical expressions for the transverse susceptibilities of single- and double-chain Ising models are written out in the Appendix. Figure 2 shows these susceptibilities as functions of temperature and demonstrates that the transverse susceptibility of a quasi-2D system only slightly deviates from that of a single chain in the entire temperature range, in contrast to the longitudinal susceptibility. Therefore, practical calculations of the transverse susceptibility of a typical quasi-2D magnet can be performed by using the formula for a single chain.

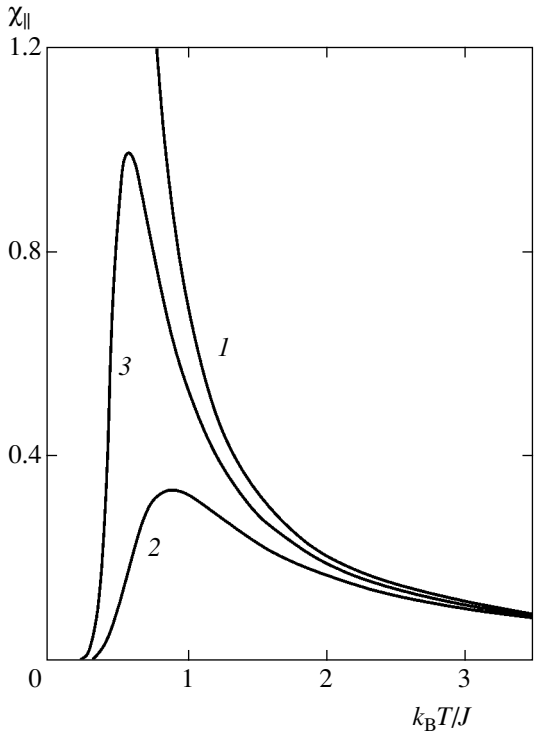


Fig. 1. Longitudinal susceptibility for Ising chain models (measured in units of $N_A g_{\parallel}^2 \mu_B^2/J$) vs. normalized temperature: (*1*) linear ferromagnetic chain; (*2*) double chain with $z'J'/J = -0.5$ ($J > 0$); (*3*) four-chain cylinder with $J > 0$ and $z'J'/2J = -0.1$.

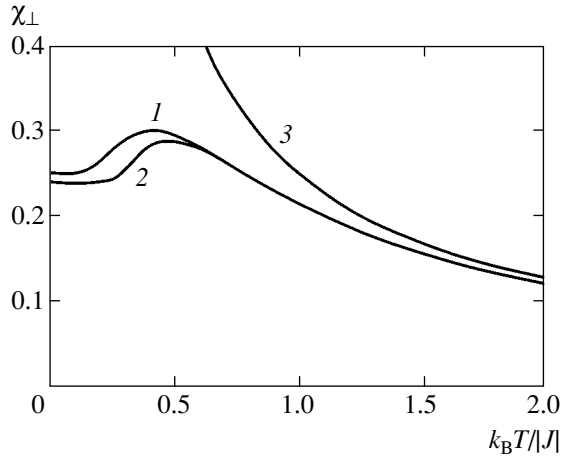


Fig. 2. Zero-field transverse susceptibility (measured in units of $N_A g_{\perp}^2 \mu_B^2/J$) vs. normalized temperature: (*1*) 1D chain ($J' = 0$); (*2*) double chain with $z'|J'/J| = 0.1$; (*3*) free spins ($J = J' = 0$).

An Ising magnet is an easy-axis one. The susceptibility of such a material in polycrystalline (powder) form is

$$\chi_p = \frac{1}{3}\chi_\parallel + \frac{2}{3}\chi_\perp. \quad (8)$$

At high temperatures, the longitudinal susceptibility of an Ising magnet obeys the Curie–Weiss law:

$$\chi_\parallel(T) \approx \frac{C_\parallel}{T-\vartheta}, \quad T \longrightarrow \infty. \quad (9)$$

This expression follows from the lowest order terms in the high-temperature expansion of the susceptibility of an infinite Ising lattice with anisotropic coupling [22, 23]. The Curie constant in (9) is expressed as

$$C_\parallel = \frac{N_A g_\parallel^2 \mu_B^2}{4k_B}, \quad (10)$$

and the Curie–Weiss temperature is

$$\vartheta = \frac{1}{k_B}\left(J + \frac{1}{2}z'J'\right). \quad (11)$$

Here, z' is the number of nearest-neighbor chains in the system: $z' = 2$ and 4 for 2D and 3D lattices, respectively.

It follows from exact low-temperature expansions obtained in [36, 38, 39] that the longitudinal susceptibility of the entire system in the two- or higher dimensional space vanishes exponentially as $T \longrightarrow 0$. At absolute zero temperature,

$$\chi_\parallel(0) = 0. \quad (12)$$

Since the clusters with even L considered here have zero magnetic moments and an infinitesimal external field cannot induce any magnetic moment that requires a finite amount of work to be done, boundary condition (12) is automatically satisfied for $L^{d-1} \times \infty$ subsystems with $d > 1$.

At high temperatures, the transverse susceptibility obeys the Curie law [36, 40]:

$$\chi_\perp(T) \approx \frac{C_\perp}{T}, \quad T \longrightarrow \infty. \quad (13)$$

(It is equal to the magnetic susceptibility of free spins.) The Curie constant in (13) is expressed as

$$C_\perp = \frac{N_A g_\perp^2 \mu_B^2}{4k_B}, \quad (14)$$

where $g_\perp$ is the transverse g factor. As $T \longrightarrow 0$, the transverse susceptibility tends to a finite limit [35, 36] (see also [40]). For the anisotropic lattices of interest for the present study,

$$\chi_\perp(0) = \frac{N_A g_\perp^2 \mu_B^2}{2(2|J| + z'|J'|)}. \quad (15)$$

An expression for the relative interchain coupling strength is found by combining (11), (14), and (15):

$$\frac{|J'|}{J} = \frac{z'}{2}\frac{C_\perp - \vartheta\chi_\perp(0)}{C_\perp + \vartheta\chi_\perp(0)}. \quad (16)$$

For $|J'|/J > 0$, it entails the constraint

$$\chi_\perp(0) < \frac{C_\perp}{\vartheta}. \quad (17)$$

Thus, the susceptibility measured at high temperatures imposes a constraint on its value at extremely low temperatures. A comparison shows that the experimental data obtained for FeTAC in [7] satisfy inequality (17).

According to (8)–(11), (13), and (14), the powder susceptibility at high temperatures also decreases as

$$\chi_p(T) \approx \frac{C}{T-\Theta} \quad (T \longrightarrow \infty), \quad (18)$$

where

$$C = \frac{N_A(g_\parallel^2 + 2g_\perp^2)\mu_B^2}{12k_B}, \quad (19)$$

$$\Theta = \frac{g_\parallel^2}{(g_\parallel^2 + 2g_\perp^2)k_B}\left(J + \frac{1}{z}z'J'\right). \quad (20)$$

At zero temperature, the powder susceptibility reduces to its transverse component:

$$\chi_p(0) = \frac{N_A g_\perp^2 \mu_B^2}{3(2|J| + z'|J'|)}. \quad (21)$$

Expressions (19)–(21) combined with the inequality $(g_\parallel^2 - 2g_\perp^2)^2 \geq 0$ yield a constraint on the relative coupling strength in an anisotropic system:

$$\left|\frac{J'}{J}\right| \leq \frac{2}{z'}\frac{C - 4\Theta\chi_p(0)}{C + 4\Theta\chi_p(0)}. \quad (22)$$

This inequality, in turn, entails an upper bound for the low-temperature plateau in powder susceptibility with $\Theta > 0$:

$$\chi_p(0) < \frac{C}{4\Theta}. \quad (23)$$

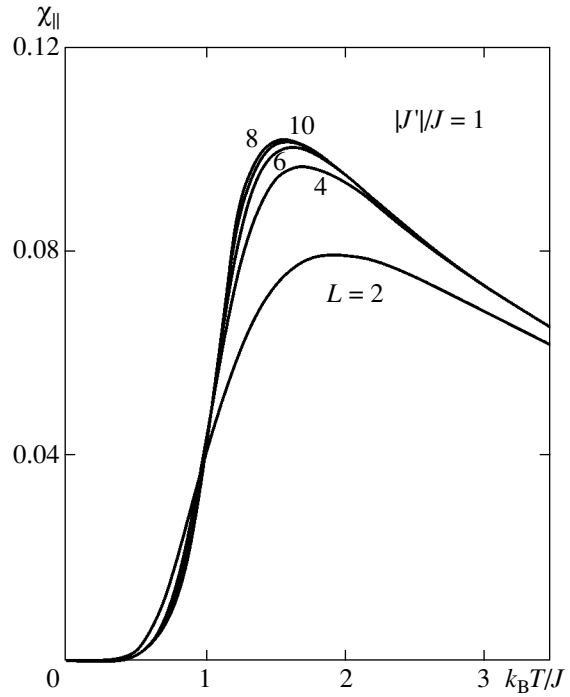


Fig. 3. Longitudinal susceptibility for $L \times \infty$ superantiferromagnetic Ising strips with L = 2–10 and $|J'|/J = 1$ (measured in units of $N_A g_\parallel^2 \mu_B^2 / J$).

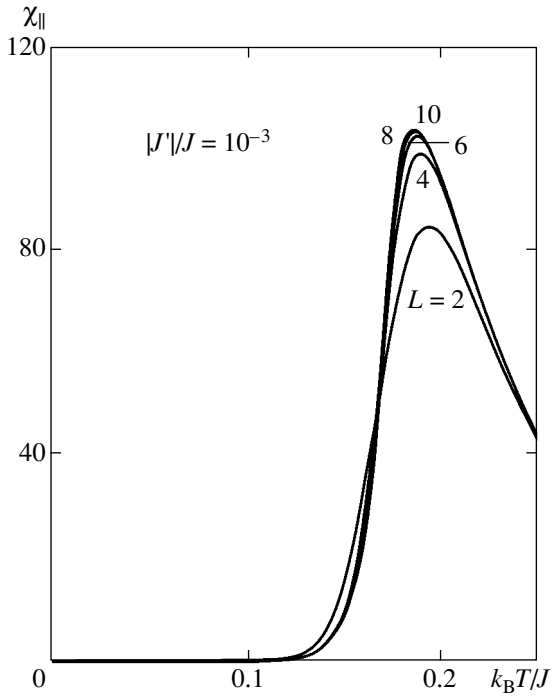


Fig. 4. Longitudinal susceptibility for $L \times \infty$ superantiferromagnetic Ising strips with L = 2–10 and $|J'|/J = 10^{-3}$ (measured in units of $N_A g_\parallel^2 \mu_B^2 / J$).

It is shown in Section 6 that available experimental data for $CoCl_2 \cdot 2NC_5H_5$ [11] support this inequality.

3. CONVERGENCE ANALYSIS OF CLUSTER EXPANSIONS FOR 2D SYSTEMS

Before applying any particular approximation to quantify experimental data, it should be verified that the systematic error of the approximation is smaller (at least, not greater) than the measurement errors.

First, let us analyze the convergence of cluster expansions for 2D systems. Figures 3 and 4 show the longitudinal susceptibilities of $L \times \infty$ Ising strips with L varying from 2 to 10 computed for $|J'|/J = 1$ and 10^{-3}, respectively. The interval between these extreme values contains the relative interchain coupling strengths characteristic of most Ising magnets actually used in experiments. When $|J'/J|$ is smaller, dipole–dipole interaction plays a significant role.

The susceptibilities were computed by using formula (5), where the eigenvalues and eigenvectors were found by direct numerical diagonalization of the starting transfer matrices having dimensions no higher than $2^{10} = 1024$ with the use of the C subroutines *tred2* and *tqli* [41]. The output data also included the coordinates of the susceptibility maximum.

Figures 3 and 4 illustrate the convergence of susceptibility with increasing subsystem size. In both extreme cases, $|J'|/J = 1$ and 10^{-3}, the curves obtained for the strips with $L = 8$ and 10 nearly coincide; i.e., these results correspond to the infinite 2D lattice up to the resolution of the graphs.

The approximation accuracy can be reliably estimated by comparing the maximum values of longitudinal susceptibility. The maximum is in the subcritical region, where the convergence follows a stretched exponential, if not a power, law.

Table 1 lists the coordinates of the longitudinal susceptibility maxima for Ising cylinders. The extrapolation to the thermodynamic limit ($L = \infty$) was performed by applying the Shanks transform [26, p. 225], which maps a sequence $\{a_l\}$ to $\{a_l'\}$ according to the formula

$$a_l' = \frac{a_{l-1}a_{l+1} - a_l^2}{a_{l-1} + a_{l+1} - 2a_l}. \tag{24}$$

The results presented in Table 1 demonstrate the following trends. The relative estimation error decreases with weakening interchain coupling from 1.3% for $|J'|/J = 1$ to 0.27% for $|J'|/J = 10^{-3}$. Furthermore, the maximum longitudinal susceptibility increases with decreasing $|J'|/J$ as

$$\frac{J\chi_\parallel^{(\max)}}{N_A g_\parallel^2 \mu_B^2} \approx \frac{0.1}{|J'|/J}. \tag{25}$$

Table 1. Coordinates of longitudinal-susceptibility maxima for cyclic $L \times \infty$ superantiferromagnetic Ising strips with different L and $\Delta = |J'|/J$ ($J > 0$, $J' < 0$): upper and lower values are $k_B T_{max}/J$ and $J\chi_{\parallel,\,max}^{(L\times\infty)}/N_A g_\parallel^2 \mu_B^2$, respectively. Extrapolation to $2D$ strips with $L = \infty$ is performed by applying Shanks transform (24) to strips with $L = 6$, 8, and 10.

L	$\Delta = 1$	$\Delta = 10^{-1}$	$\Delta = 10^{-2}$	$\Delta = 10^{-3}$
2	1.957443	0.593380	0.301827	0.193399
	0.080142	0.853605	8.577071	85.786055
4	1.723455	0.563281	0.292302	0.189102
	0.097244	1.000137	10.019720	100.203811
6	1.640401	0.550014	0.287992	0.187135
	0.101008	1.034448	10.360714	103.612905
8	1.603729	0.543665	0.285905	0.186177
	0.102075	1.044801	10.463915	104.644760
10	1.587364	0.540669	0.284914	0.185721
	0.102404	1.048171	10.497584	104.981419
∞	1.57(2)	0.538(3)	0.2840(9)	0.1853(5)
	0.1026(2)	1.050(2)	10.51(2)	105.1(2)

However, the error of estimation of the maximum value, unlike $k_B T_{max}/J$, is almost independent of lattice anisotropy. For $L = 10$, it is approximately 0.2%. This accuracy is sufficient for quantitative description of available experimental data.

Let us now discuss the accuracy of the approximation used in [18]. According to Table 1, we find that the maximum longitudinal susceptibility of the 2D lattice with $|J'|/J = 10^{-1}$ corresponds to $k_B T_{max}/J = 0.538(3)$, and its value is $J\chi_\parallel^{(max)}/N_A g_\parallel^2 \mu_B^2 = 1.050(2)$. On the other hand, the results presented in [18, Fig. 5] for the same superantiferromagnet demonstrate that the maximum has the coordinates $k_B T_{max}/J \approx 0.656$ and $J\chi_\parallel^{(max)} \approx 0.809$ (with redefined coupling constants). Thus, the errors of estimation of the maximum value and the corresponding temperature in the approximation used in [18] are 22 and 23%, respectively; i.e., the theory developed in [18] is too inaccurate to be applicable to experimental data even for weakly anisotropic lattices.

The inverse variation of maximum longitudinal susceptibility of superantiferromagnetic lattices with relative interchain coupling strength described by empirical formula (25) can be explained as follows. The critical temperature for anisotropic 2D Ising lattice satisfies the equation [42]

$$\sinh\frac{|J|}{k_B T_c}\sinh\frac{|J'|}{k_B T_c} = 1. \qquad (26)$$

As $|J'|/J \longrightarrow 0$, this equation yields

$$\frac{|J'|}{J} \approx 2\frac{k_B T_c}{J}\exp\left(-\frac{J}{k_B T_c}\right). \qquad (27)$$

On the other hand, the longitudinal susceptibility of a typical quasi-1D Ising superantiferromagnet at temperatures slightly above the maximum point is well approximated by the formula for the longitudinal susceptibility of a single Ising chain. Moreover, the maximum point approaches the critical temperature with increasing lattice anisotropy. Therefore, assuming that $T_{max} \approx T_c$ and $\chi_\parallel^{(max)} \propto \chi_\parallel^{(1D)}(T_{max})$, we combine (27) with (A.1) to obtain

$$\frac{J\chi_\parallel^{(max)}}{N_A g_\parallel^2 \mu_B^2} \propto \left(\frac{|J'|}{J}\right)^{-1}. \qquad (28)$$

The data listed in Table 1 demonstrate that relation (28) holds in a surprisingly wide interval extending almost to $|J'| = J$.

4. ZERO-FIELD MAGNETIC SUSCEPTIBILITY OF SINGLE-CRYSTAL FeTAC

Single crystals of ferrous trimethylammonium chloride (FeTAC) are characterized by the most pronounced quasi-one-dimensional magnetic ordering among the class of compounds described by the formula $[(CH_3)_3NH]MX_3 \cdot 2H_2O$, where M denotes a metal (such as Co, Fe, or Ni) and X is chlorine or bromine (see [8, 43] and references therein). Physical properties of FeTAC are the subject of extensive experimental studies.

The static zero-field magnetic susceptibility of FeTAC single crystals was measured in [7] at temperatures ranging from 1.4 to 300 K. The susceptibility along the easy axis (crystallographic b axis) is interpreted as the longitudinal susceptibility: $\chi_b \equiv \chi_\parallel$. According to [7], its maximum value $\chi_b^{(max)} = 100$ cm^3/mol is reached at $T_{max} = 3.18(2)$ K, and the critical temperature determined from the steepest slope of longitudinal susceptibility below T_{max} is $T_c = 3.12(2)$ K (i.e., $T_{max}/T_c \approx 1.02$). At temperatures well above T_{max}, the susceptibility obeys the Curie–Weiss law.

The quantitative interpretation of measured susceptibilities presented in [7] is based on the single-chain approximation. Therefore, it is applicable only at temperatures above T_{max}. By fitting the longitudinal susceptibility of the 1D Ising chain to experimental data points in the interval between 6 and 18 K, it was found that $C_b = C_\parallel = 5.52(4)$ cm^3 K/mol (Curie constant) and $\vartheta \approx J/k_B = 16.6(1)$ K [7]. Combining these results with (10), we obtain the longitudinal g factor: $g_\parallel = 7.67(3)$.

Moreover, the relative interchain coupling strength was estimated in [7] by using a well-known expression for the susceptibility of a quasi-1D system with coupling between chains described in the molecular field approximation [44],

$$\chi_{\parallel}^{(\mathrm{MFA})} = \chi_{\parallel}^{(1\mathrm{D})} \Big/ \left(1 - \frac{z'J'}{2C_{\parallel}}\chi_{\parallel}^{(1\mathrm{D})}\right). \qquad (29)$$

By fitting this theoretical formula to experimental data on susceptibility in the interval between 3.2 K (again above $T_{\max}$) and 18 K, the approximate value $J'/J \approx -2 \times 10^{-3}$ was obtained [7].

According to the results of measurements of the specific heat of FeTAC reported in [8], $T_c = 3.125(5)$ K and $J/k_B = 17.7(3)$ K. Onsager's solution (26) was used in [8] to obtain $|J'/J| = 1.3 \times 10^{-3}$ for FeTAC.

Before discussing the results on FeTAC obtained in the present study, let us note that the use of dependence of k_BT_c/J on J'/J is not the best method for finding the interchain coupling strength in quasi-1D systems. Indeed, (27) entails the following relation between the relative errors in normalized interchain coupling and reduced temperature:

$$\delta\left(\frac{|J'|}{J}\right) = \left[1 + \left(\frac{k_BT_c}{J}\right)^{-1}\right]\delta\left(\frac{k_BT_c}{J}\right). \qquad (30)$$

Therefore, the error in $|J'|/J$ estimated from T_c increases with coupling anisotropy (since k_BT_c/J decreases).

The same conclusion can be reached in a different manner. Transcendental equation (26) is easily solved on a computer to obtain curve *1* in Fig. 5. However, the inaccuracy of input data should be taken into account. Following [8], let us use $T_c = 3.125 \pm 0.005$ K and $J/k_B = 17.7 \pm 0.3$ K. Then, $k_BT_c/J = 0.177 \pm 0.003$, the corresponding relative error is 1.7%, and the relative error in the result $|J'|/J = (1.1\text{–}1.4) \times 10^{-3}$ obtained by solving (26) is 12%. This sharp increase in error is explained by a rapid increase in the steepness of curve *1* with decreasing J'/J (see Fig. 5).

Alternatively, the ratio $|J'|/J$ can be determined for a superferromagnetic system by using maximum susceptibility values (curve *2* in Fig. 5). According to (25), $J\chi_{\parallel}^{(\max)}/N_Ag_{\parallel}^2\mu_B^2$ varies in inverse proportion to $|J'|/J$. This relation obviously implies that the respective relative errors in these parameters are equal; therefore,

$$\delta\left(\frac{|J'|}{J}\right) = \delta(\chi_{\parallel}^{(\max)}) + \delta\left(\frac{J}{k_B}\right) + \delta(C_{\parallel}). \qquad (31)$$

One important advantage of this method for estimating $|J'|/J$ over the one discussed above is that the error is independent of lattice anisotropy.

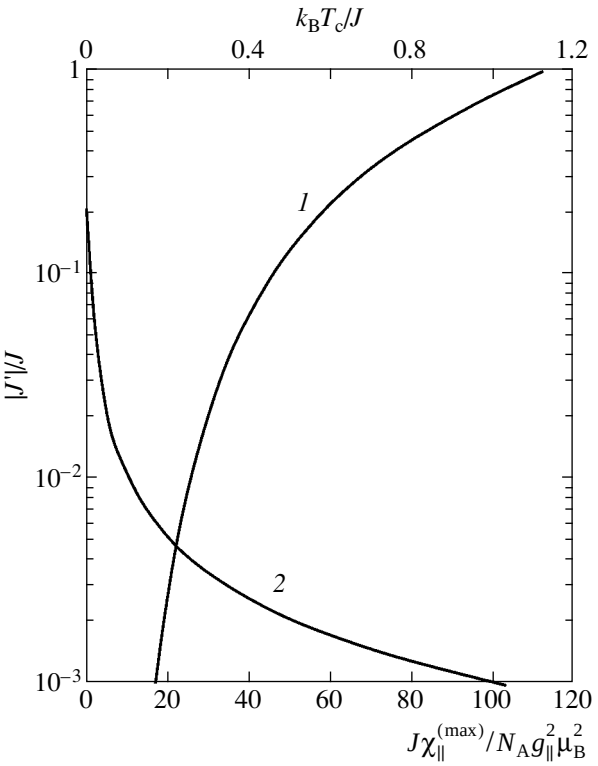


Fig. 5. Reduced critical temperature (curve *1*) and maximum value of normalized longitudinal susceptibility (curve *2*) vs. $|J'|/J$ for 2D superantiferromagnetic Ising lattices.

The principal result of this section is a quantitative description of the magnetic susceptibility of FeTAC in the entire temperature range down to absolute zero. Computations were performed for a 10-chain strip to ensure that the approximation error is much smaller than measurement errors. ($L \times \infty$ strips with $L = 12$, 14, and larger could easily be simulated on a modern computer if necessary.)

Expression (5) for longitudinal susceptibility contains three parameters: J/k_B, $g_{\parallel}$, and J'/J. As in [7], the function $\chi_{\parallel}^{(1\mathrm{D})}(T)$ was fitted to experimental data points at temperatures of 6 K and higher to obtain $J/k_B = 16.6$ K and $g_{\parallel} = 7.67$. By matching the computed maximum susceptibility with that measured in [7] ($T_{\max} = 3.18$ K, $\chi_{\parallel}^{(\max)} = 100$ cm^3/mol), it was found that $J'/J = -0.00138$ with an error in the last digit.

Figure 6 demonstrates that the curve computed in the present study agrees with experimental data for FeTAC obtained in [7]. It should be reiterated here that the theoretical description of experimental data on susceptibility given in [7] is valid only at temperatures above the maximum point, when the single-chain approximation is applicable.

5. CLUSTER EXPANSIONS FOR 3D SYSTEMS

Modern computers can be used to simulate 3D Ising chain clusters only for $L \le 4$. Note that exact expressions are available for the susceptibilities of the double-chain cluster and the $2 \times 2 \times \infty$ parallelepiped (see Appendix).

To find the eigenvalues of the 65536×65536 transfer matrix of the cyclic $4 \times 4 \times \infty$ Ising parallelepiped, both lattice and spin symmetries were used to represent it in a block diagonal form consisting of 433×433 and 372×372 subblocks, whose eigenvalues and eigenvectors are required to calculate the susceptibility given by (5) [34]. Exact diagonalization of these relatively small subblocks can readily be performed on a PC.

The next larger cluster that should have been simulated in the present study is the $6 \times 6 \times \infty$ parallelepiped. However, the corresponding transfer matrix is $2^{36} \times 2^{36}$, and the dimensions of the subblocks in its block diagonal form determined by using the symmetries of the system (as in the case of the $4 \times 4 \times \infty$ cluster) are 119583470 and 119539680 [34]. The complete solution of the spectral problem for these matrices is far beyond the capabilities of present-day supercomputers. To date, computations have been performed for the $6 \times 6 \times \infty$ Ising system only at the quantum limit and a few lowest eigenvalues of the corresponding sparse Hamiltonian matrix have been calculated [45].

For this reason, the present analysis is restricted to $2 \times \infty$, $2 \times 2 \times \infty$, and $4 \times 4 \times \infty$ clusters. Figure 7 illustrates the convergence of the zero-field susceptibilities computed for 3D clusters (curves *2–4*) with respect to the number of chains in a cluster. The trends shown here are qualitatively similar to those manifested in the 2D simulations (Figs. 3 and 4): the maximum susceptibility increases with the cluster size, while the corresponding reduced temperature decreases. As $L \longrightarrow \infty$, the coordinates of the maximum must approach their respective limits. However, these limits cannot be calculated by Shanks extrapolation for lack of solution for the $6 \times 6 \times \infty$ cluster. (Unfortunately, the $2 \times \infty$ chain cannot be used in an extrapolation process, because it is not a truly 3D cluster.)

The accuracy of the $4 \times 4 \times \infty$ approximation used here for comparison with experiment can be estimated indirectly by invoking the results of a Padé–Borel resummation of high-temperature expansions for the susceptibility of an anisotropic 3D Ising lattice [24]. In that study, the coordinates of the superantiferromagnetic susceptibility maximum were presented as func-

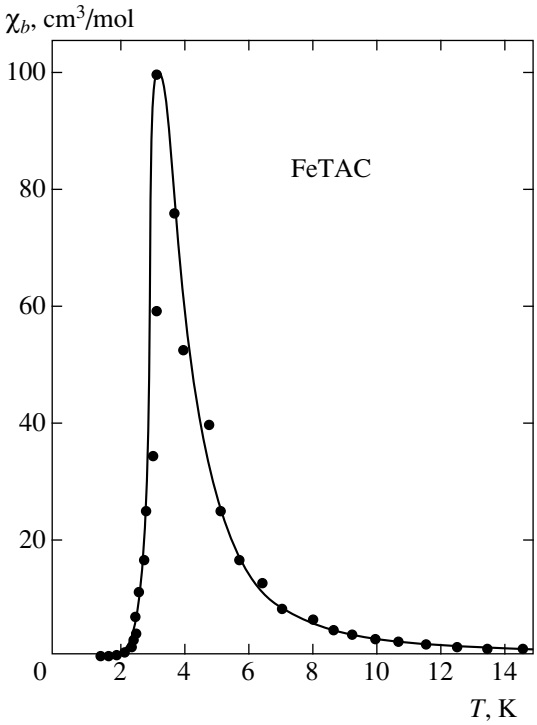


Fig. 6. Magnetic susceptibility of FeTAC along the crystallographic *b* axis: measurement data from [7] (symbols) and $\chi_\parallel^{(10\times\infty)}(T)$ calculated for $J/k_B = 16.6$ K, $g_\parallel = 7.67$, and $J'/J = -1.38 \times 10^{-3}$ (curve).

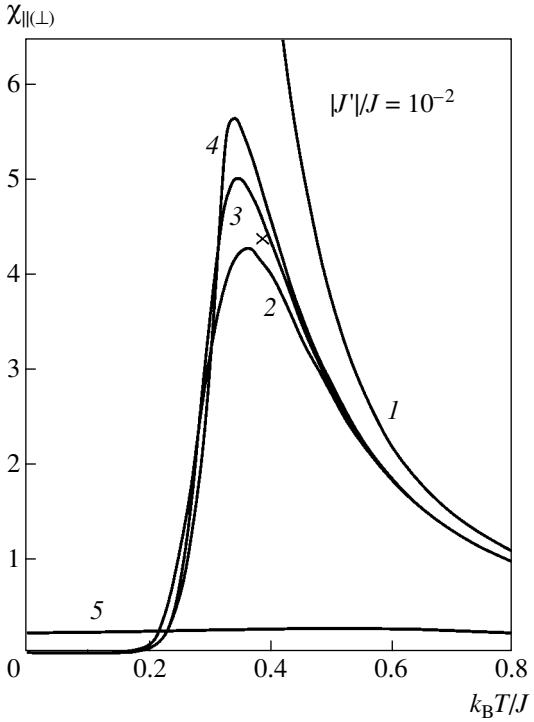


Fig. 7. Longitudinal susceptibility for Ising clusters with $|J'|/J = 10^{-2}$ (measured in units of $N_A g_\parallel^2 \mu_B^2/J$): (*1*) linear ferromagnetic chain; (*2*) double chain with quadrupled interchain coupling; (*3*) cyclic $2 \times 2 \times \infty$ and (*4*) cyclic $4 \times 4 \times \infty$ parallelepipeds. The cross is the maximum calculated in [24]. The transverse susceptibility of a linear Ising chain (measured in units of $N_A g_\perp^2 \mu_B^2/J$) is shown for comparison (curve *5*).

tions of J'/J (see Table 1 in [24]). In the isotropic case ($|J'|/J = 1$), when high-temperature expansions must lead to reliable results, it was found that $k_B T_{max}/J = 2.400(25)$ and $J\chi_\parallel^{(max)}/N_A g_\parallel^2 \mu_B^2 = 0.058$ [24]. In this study, $J\chi_{\parallel,\,max}^{(4\times4\times\infty)}/N_A g_\parallel^2 \mu_B^2 = 0.05547$ is obtained for the $4 \times 4 \times \infty$ lattice, which is lower by 4.35%. By analogy with the 2D simulations, it can be assumed that the relative error in the maximum value is independent of lattice anisotropy. (Actually, the error even slightly decreases with $|J'|/J$, which can easily be demonstrated by using the results for $L = 4$ presented in [24, Table 1].) Then, the error of the $4 \times 4 \times \infty$ approximation must be about 4%. Since theoretical results are compared here with experimental data on powder susceptibility, which are not very accurate, the $4 \times 4 \times \infty$ cluster approximation is well suited for such a comparison.

As mentioned in the Introduction, high-temperature expansions lead to increasingly inaccurate results with decreasing $|J'/J|$, because they contain a small number of terms. The cross in Fig. 7 represents the susceptibility maximum calculated in [24] for $|J'|/J = 10^{-2}$, with $k_B T/J = 0.385(5)$ and $J\chi_\parallel/N_A g_\parallel^2 \mu_B^2 = 4.40(75)$. According to the figure, high-temperature expansions are even less reliable than the $2 \times 2 \times \infty$ approximation for this degree of anisotropy. These results cannot be compared to experimental data for such highly anisotropic superantiferromagnets as $CoCl_2 \cdot 2NC_5H_5$ (with $|J'|/J < 10^{-2}$). Now, it is clear that the formal quantitative comparison of this kind presented in [24] is groundless.

Table 2 summarizes the coordinates of the susceptibility maximum calculated for the $4 \times 4 \times \infty$ superantiferromagnetic cluster. (The maximum longitudinal susceptibility is divided by 3 with a view to comparing with powder susceptibility.) As in the case of 2D lattice, the maximum longitudinal susceptibility varies in inverse proportion with $|J'|/J$:

$$\frac{J\chi_{\parallel,\,max}^{(4\times4\times\infty)}}{3N_A g_\parallel^2 \mu_B^2} \approx \frac{0.019}{|J'|/J}. \quad (32)$$

This behavior is explained by analogy with the 2D case: when the coupling between chains is described in the molecular field approximation (while the intrachain coupling is modeled exactly), the critical temperature for the ferromagnetic simple cubic Ising lattice satisfies the transcendental equation [44]

$$K_B T_c = \frac{z'}{2} J' \exp\left(\frac{J}{k_B T_c}\right). \quad (33)$$

Since $T_{max} \approx T_c$ and the maximum susceptibility is on the order of $\chi_\parallel^{(1D)}(T_{max})$ for anisotropic superantiferromagnetic lattices with sufficiently high anisotropy, the law formulated in (28) follows again from Eq. (A.1).

Table 2. Normalized coordinates of longitudinal-susceptibility maximum as a function of J'/J for $4 \times 4 \times \infty$ lattices with $J > 0$ and $J' < 0$

J'/J	$k_B T_{max}/J$	$\frac{1}{3}\chi_{\parallel,\,max}^{(4\times4\times\infty)}$
−1.0000	2.5833625	0.01849229
−0.1000	0.7155626	0.18871242
−0.0100	0.3399309	1.89015328
−0.0095	0.3346111	1.98964522
−0.0090	0.3301889	2.10018741
−0.0085	0.3254978	2.22374800
−0.0080	0.3207257	2.36273906
−0.0075	0.3158513	2.52025948
−0.0070	0.3105331	2.70029750
−0.0065	0.3052407	2.90802557
−0.0060	0.2996506	3.15038075
−0.0055	0.2936108	3.43680225
−0.0050	0.2875449	3.78048335
−0.0045	0.2808194	4.20055926
−0.0040	0.2738453	4.72563875
−0.0035	0.2661257	5.40079611
−0.0030	0.2578018	6.30088685
−0.0025	0.2483499	7.56111673
−0.0020	0.2378198	9.45146198
−0.0015	0.2252555	12.6019164
−0.0010	0.2095811	18.9030324

Table 3 lists the coordinates of the inflection point below the maximum and the corresponding slopes of the susceptibility curve. It is clear that the abscissa of the inflection point is a lower bound for critical temperature (cf. [34]). The numerical results presented in Table 3 demonstrate a rapid increase in the normalized $\chi' = \partial\chi/\partial T$ with decreasing $|J'|/J$ and show that the susceptibility at the critical point also varies in inverse proportion to $|J'|/J$.

6. POWDER MAGNETIC SUSCEPTIBILITY OF $CoCl_2 \cdot 2NC_5H_5$ AND $FeCl_2 \cdot 2NC_5H_5$

The spin systems of crystals of the pyridine complexes of cobalt and iron(II) chlorides are 3D Ising lattices with quasi-1D coupling [1, 2]. In these compounds, Co^{2+} or Fe^{2+} ions are linked by next-nearest-neighbor superexchange coupling through chlorine into linear chains separated by pyridine rings, which are responsible for interchain coupling. At temperatures on

Table 3. Normalized coordinates of the inflection point of longitudinal susceptibility below its maximum and the values of susceptibility and its temperature derivative at the inflection point as functions of J'/J for $4 \times 4 \times \infty$ lattices with $J > 0$ and $J' < 0$

J'/J	$k_B T_c/J$	$\frac{1}{3}\chi'^{(4\times4\times\infty)}_{\parallel,\,c}$	$\frac{1}{3}\chi^{(4\times4\times\infty)}_{\parallel,\,c}$
−1.0000	1.947425	0.06298528	0.01244998
−0.1000	0.621899	3.41073780	0.13753748
−0.0100	0.311571	110.197410	1.39414706
−0.0095	0.307526	119.120733	1.45608732
−0.0090	0.303649	127.885372	1.53655666
−0.0085	0.299877	139.597874	1.63723231
−0.0080	0.296189	151.066706	1.76240864
−0.0075	0.291462	166.471661	1.85697964
−0.0070	0.287312	182.003965	2.00942967
−0.0065	0.281839	205.372014	2.11081321
−0.0060	0.277612	227.136261	2.33597593
−0.0055	0.272541	254.505992	2.55608624
−0.0050	0.267087	290.445604	2.81257312
−0.0045	0.260747	336.356921	3.06639337
−0.0040	0.254937	394.237723	3.50292699
−0.0035	0.248861	473.481579	4.10957821
−0.0030	0.241041	591.045463	4.72049719
−0.0025	0.232527	745.165859	5.60393984
−0.0020	0.222934	1017.23093	6.95010013
−0.0015	0.211715	1510.16621	9.23674430
−0.0010	0.197356	2567.00852	13.6273345

the order of the critical temperature, the only significantly populated level in the energy level system of metal ions modified by the single-ion anisotropy field is the ground-state Kramers doublet, which is separated from higher levels by a large energy gap. Therefore, the coupling between these ions can be modeled by an effective spin-1/2 Hamiltonian.

The data points in Fig. 8 represent the zero-field magnetic susceptibilities of polycrystalline $CoCl_2 \cdot 2NC_5H_5$ and $FeCl_2 \cdot 2NC_5H_5$ powders measured in [11] and [17], respectively.

For $CoCl_2 \cdot 2NC_5H_5$ crystals, the Curie constant is $C = 2.82(5)$ cm^3/mol, the Curie–Weiss temperature is $\Theta = 4.95(5)$ K, and the low-temperature limit value of powder susceptibility is $\chi_p(0) = 0.14(1)$ cm^3/mol [11]. These numerical values are consistent with upper bound (23). Unfortunately, inequality (22) cannot be used to obtain any useful quantitative information because of a large experimental error. However, large values of maximum susceptibility (see Fig. 8) suggest that the interchain coupling is weak. The corresponding contribution to longitudinal susceptibility can be ignored, since it is much smaller than the $\chi_p(0)$ measurement error. If the contribution of J' is also ignored, then the decoupled system (19)–(21) yields $J/k_B = 11.4$ K, $g_\parallel = 6.26$, and $g_\perp = 5.05$, and only the value of J'/J is required to calculate the susceptibility curve.

The critical temperature and maximum susceptibility calculated for the 3D Ising lattice as functions of $|J'|/J$ are qualitatively similar to those shown in Fig. 5 for the 2D model. Numerical values of $k_B T_c/J$ in the 3D Ising model can be found in a recent paper ([34], Table 3, the values in the T_c column divided by 2). Using these values and taking $T_c = 3.17(2)$ K and $J/k_B = 10.6(6)$ K for $CoCl_2 \cdot 2NC_5H_5$ from [11], we obtain $|J'|/J = 0.0069^{+0.0022}_{-0.0016}$ with an error of 30%. Analogously, taking $T_c = 6.6(3)$ K and $J/k_B = 25(2)$ K for $FeCl_2 \cdot 2NC_5H_5$ from [17], we obtain $|J'|/J = 0.0038^{+0.0033}_{-0.0015}$; i.e., the corresponding error is even larger. Therefore, $|J'|/J$ should again be determined from the maximum susceptibility.

Figure 7 demonstrates that the transverse susceptibility in the neighborhood of the longitudinal-susceptibility maximum point is almost constant, and its value predicted by (15) combined with (A.12) is $[\chi_\perp^{(1D)}(0) + \chi_{\perp,\,\max}^{(1D)}]/2 \approx 0.27496$ (measured in units of $J/N_A g_\parallel^2 \mu_B^2$). Accordingly, the maximum powder susceptibility of quasi-1D superantiferromagnets with $10^{-3} \le |J'|/J \le 10^{-2}$ can be calculated as

$$\frac{J\chi_p^{(\max)}}{N_A g_\parallel^2 \mu_B^2} \approx \frac{1}{3}\frac{J\chi_{\parallel,\,\max}^{(4\times4\times\infty)}}{N_A g_\parallel^2 \mu_B^2} + 0.1833\left(\frac{g_\perp}{g_\parallel}\right)^2, \qquad (34)$$

with $J\chi_{\parallel,\,\max}^{(4\times4\times\infty)}/3N_A g_\parallel^2 \mu_B^2$ taken from Table 2.

Using the experimental value $\chi_p^{(\max)} = 3.9$ cm^3/mol for $CoCl_2 \cdot 2NC_5H_5$ determined from Fig. 4 in [11] and applying the $4 \times 4 \times \infty$ cluster model, we find that $|J'|/J = 6.53 \times 10^{-3}$.

Curve *1* in Fig. 8 is the powder susceptibility

$$\chi_p(T) \approx \frac{1}{3}[\chi_\parallel^{(4\times4\times\infty)}(T) + 2\chi_\perp^{(1D)}(T)] \qquad (35)$$

calculated by using the parameters obtained for $CoCl_2 \cdot 2NC_5H_5$. Here, the maximum is located at 3.48 K, which agrees with $T_{max} = 3.51(1)$ K measured in [11]. The inflection point below the maximum of the theoretical curve is located at 3.2 K. This result is also consistent with $T_c = 3.17(2)$ K determined by measuring specific heat in [11].

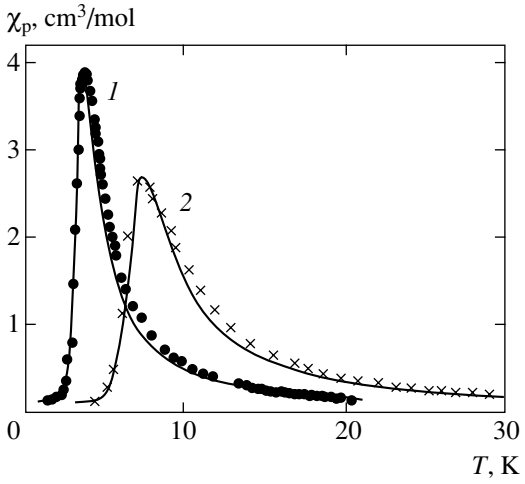


Fig. 8. Temperature dependence of $CoCl_2 \cdot 2NC_5H_5$ (*1*) and $FeCl_2 \cdot 2NC_5H_5$ (*2*) powder susceptibility: measurement data from [11] (circles) and [17] (crosses) and theoretical predictions based on (35) with $J/k_B = 11.4$ K, $J'/J = -6.53 \times 10^{-3}$, $g_\parallel = 6.26$, and $g_\perp = 5.05$ for $CoCl_2 \cdot 2NC_5H_5$ and $J/k_B = 24.5$ K, $J'/J = -5.65 \times 10^{-3}$, $g_\parallel = 7.07$, and $g_\perp = 6.85$ for $FeCl_2 \cdot 2NC_5H_5$ (curves).

The experimental data for $FeCl_2 \cdot 2NC_5H_5$ presented in [17] were used to obtain $C = 4.5$ cm^3/mol, $\Theta = 8.5$ K, $\chi_p(0) = 0.12$ cm^3/mol, and $\chi_p^{(max)} = 2.7$ cm^3/mol. Again, the model parameters were adjusted to find $J/k_B = 24.5$ K, $|J'|/J = 5.65 \times 10^{-3}$, $g_\parallel = 7.07$, and $g_\perp = 6.85$. These parameters were used in (35) to obtain curve *2* in Fig. 8, for which $T_{max} = 7.2$ K and $T_c = 6.7$ K. According to [17], $J/k_B = 25(2)$ and $T_c = 6.6(3)$; i.e., the estimates for J/k_B and T_c obtained in this study agree with experimental data within measurement error.

7. CONCLUSIONS

The temperature dependence of the longitudinal and transverse zero-field susceptibilities of 2D and 3D Ising lattices with anisotropic coupling is analyzed. The analysis is based on approximations of the original lattices with ensembles of independent chain clusters that are infinitely long in the strong-coupling direction.

A detailed treatment is presented of the ferromagnetic intrachain and antiferromagnetic interchain couplings that constitute the superantiferromagnetic coupling characteristic of the modeled systems. For this coupling configuration, longitudinal susceptibility has a maximum whose value varies in inverse proportion to interchain coupling strength. An explanation is proposed for the inverse proportionality.

It is found that the relative error in the value of the maximum calculated for clusters of the same finite size L is independent of lattice anisotropy.

For highly anisotropic superantiferromagnets, it is shown that the interchain coupling strength can be determined much more accurately by using the maximum value than the critical temperature, whereas the latter method is applicable to weakly anisotropic systems.

A convergence analysis of cluster expansions performed for 2D systems shows that $L \times \infty$ strips of width $L = 10$ can be used to calculate the maximum susceptibility up to an error of 0.2%. If required, clusters of larger size can be used to improve accuracy. Currently, strips of width $L \le 10$ can be simulated on a standard PC. Simulations for widths up to $L = 16$ can be performed on supercomputers. Strips of larger size can be simulated by using the symmetry of $L \times \infty$ cylinders and representing the transfer matrix in block diagonal form.

Three-dimensional simulation is an essentially different task. The susceptibility of $L \times L \times \infty$ Ising systems with $L \le 4$ can be calculated on a PC, whereas the $6 \times 6 \times \infty$ problem cannot be solved on any supercomputer even after the transfer matrix is reduced to a block diagonal form.

The approximation accuracy achieved in this study is sufficient for a well-founded quantitative description of the magnetic susceptibilities measured for real 2D and 3D anisotropic Ising superantiferromagnets. The present numerical results obtained are valid in the entire experimental temperature range.

The agreement achieved between theory and experiment strongly suggests that FeTAC, $CoCl_2 \cdot 2NC_5H_5$, and $FeCl_2 \cdot 2NC_5H_5$ can be very accurately treated as Ising magnets.

The accuracy of estimation of $|J'|/J$ is improved from one or two to three digits.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project nos. 03-02-16909 and 04-03-32528.

APPENDIX

The formulas given below were used in analyzing experimental data.

Longitudinal Susceptibility

The zero-field longitudinal susceptibility of a 1D Ising chain is

$$\chi_\parallel^{(1D)}(T) = \frac{N_A g_\parallel^2 \mu_B^2}{4k_B T} e^{J/k_B T}. \quad (A.1)$$

The longitudinal susceptibility of a double-chain 1D ($2 \cdot 1D$) Ising model is expressed as follows [46, 47]

(see also [48]):

$$\chi_{\parallel}^{(2\cdot 1D)}(T) = \frac{N_A g_{\parallel}^2 \mu_B^2}{4k_B T} \frac{e^{J/k_B T}}{A\cosh(z'J'/2k_B T)} \times\left(A + \cosh\frac{J}{k_B T}\sinh\frac{z'J'}{2k_B T}\right) \times\left(A + \sinh\frac{J}{k_B T}\sin\frac{z'J'}{2k_B T}\right), \quad (A.2)$$

where

$$A = \left(1 + \cosh^2\frac{J}{k_B T}\sinh^2\frac{z'J'}{2k_B T}\right)^{1/2}. \quad (A.3)$$

The expression for the longitudinal susceptibility of a four-chain (4 · 1D) Ising model (a truly 3D cluster, such as $4 \times \infty$ cylinder or $2 \times 2 \times \infty$ parallelepiped) is [30]

$$\chi_{\parallel}^{(4\cdot 1D)}(T) = \frac{N_A g_{\parallel}^2 \mu_B^2}{16k_B T} \times\frac{(A^2 - 4e^{2y}\sinh^4 x)F + 4AGe^y\sinh x}{(A^2 - 2Ae^y\sinh 2x\cosh y + 4e^{2y}\sinh^4 x)R_1R_2}, \quad (A.4)$$

where

$$R_{1,2} = [1 + (\sqrt{2}\cosh x\sinh y \pm \cosh y)^2]^{1/2},$$

$$A = (\sqrt{2}\cosh x\cosh y + \sinh y + R_1) \times(\sqrt{2}\cosh x\cosh y - \sinh y + R_2),$$

$$B = 4[1 + (\sqrt{2}\cosh x\sinh y + \cosh y + R_1) \times(\sqrt{2}\cosh x\sinh y - \cosh y + R_2)], \quad (A.5)$$

$$F = \cosh^2 y - 2\cosh^2 x\sinh^2 y + R_1R_2 + B - 3,$$

$$G = 2\sqrt{2}(R_1 + R_2) + \cosh x\sinh y(2B - F),$$

with $x = J/k_B T$ and $y = z'J'/2k_B T$.

According to these formulas, the longitudinal susceptibility either increases indefinitely or vanishes (exponentially) as $T \longrightarrow 0$, depending on whether the zero-temperature ordered state has nonzero or zero magnetic moment. In particular, if $J > 0$ and $J' < 0$, then the longitudinal susceptibility vanishes at $T = 0$. As $T \longrightarrow \infty$, the susceptibility vanishes according to the Curie–Weiss law (9). It should be noted here that the constant parameters in the law are the correct Curie constant and Curie–Weiss temperature given by (10) and (11), respectively.

Transverse Susceptibility

The zero-field transverse susceptibility of a linear Ising chain is [36, 49, 50]

$$\chi_{\perp}^{(1D)}(T) = \frac{N_A g_{\perp}^2 \mu_B^2}{4J} \times\left[\tanh\frac{J}{2k_B T} + \frac{J/2k_B T}{\cosh^2(J/2k_B T)}\right]. \quad (A.6)$$

The zero-field transverse susceptibility of a double-chain Ising model with interchain coupling strength $z'J'$ has the form [51]

$$\chi_{\perp}^{(2\cdot 1D)}(T) = \frac{1}{16}N_A g_{\perp}^2 \mu_B^2(A_1 - B_1G_1 + 2A_2G_2 - 2B_2G_3 + A_3G_4 - B_3G_5), \quad (A.7)$$

where

$$A_1 = \frac{1}{2J + z'J'}\sinh\frac{2J + z'J'}{k_B T} + \frac{1}{2J - z'J'} \times\sinh\frac{2J - z'J'}{k_B T} + \frac{2}{z'J'}\sinh\frac{z'J'}{k_B T}, \quad (A.8)$$

$$A_2 = \frac{1}{2J + z'J'}\sinh\frac{2J + z'J'}{k_B T} - \frac{1}{2J - z'J'}\sinh\frac{2J - z'J'}{k_B T},$$

$$A_3 = \frac{4}{2J + z'J'}\sinh\frac{2J + z'J'}{k_B T} - A_1 - 2A_2,$$

$$B_1 = \frac{1}{2J + z'J'}\left(\cosh\frac{2J + z'J'}{k_B T} - 1\right) - \frac{1}{2J - z'J'}\left(\cosh\frac{2J - z'J'}{k_B T} - 1\right) + \frac{2}{z'J'}\left(\cosh\frac{z'J'}{k_B T} - 1\right),$$

$$B_2 = \frac{1}{2J + z'J'}\left(\cosh\frac{2J + z'J'}{k_B T} - 1\right) + \frac{1}{2J - z'J'}\left(\cosh\frac{2J - z'J'}{k_B T} - 1\right), \quad (A.9)$$

$$B_3 = \frac{4}{2J + z'J'}\left(\cosh\frac{2J + z'J'}{k_B T} - 1\right) - B_1 - 2B_2;$$

$$G_1 = \frac{1}{R}\cosh\frac{J}{k_B T}\sinh\frac{z'J'}{2k_B T},$$

$$G_2 = \frac{1}{R}\sinh\frac{J}{k_B T}\sinh\frac{z'J'}{2k_B T},$$

$$G_3 = \frac{1}{R}\sinh\frac{J}{k_BT}\left(\cosh\frac{z'J'}{2k_BT} - \frac{1}{S}\cosh\frac{J}{k_BT}\right),$$
$$G_4 = \frac{1}{S}\sinh^2\frac{J}{k_BT}\left(\frac{2}{R}\sinh^2\frac{z'J'}{2k_BT} + \frac{1}{S}\right),$$
$$G_5 = \frac{1}{RS}\sinh^2\frac{J}{k_BT}\sinh\frac{z'J'}{2k_BT} \times\left(2\cosh\frac{z'J'}{2k_BT} - \frac{1}{S}\cosh\frac{J}{k_BT}\right), \quad \text{(A.10)}$$

with

$$R = \left(1 + \cosh^2\frac{J}{k_BT}\sinh^2\frac{z'J'}{2k_BT}\right)^{1/2},$$
$$S = R + \cosh\frac{J}{k_BT}\cosh\frac{z'J'}{2k_BT}. \quad \text{(A.11)}$$

The transverse susceptibility is invariant under the sign changes $J \longrightarrow -J$ and $J' \longrightarrow -J'$. At absolute zero temperature, susceptibility (A.7) reduces to (15), which is the correct value for a system of any dimensionality. Figure 2 demonstrates that the transverse susceptibility is nearly constant at low temperatures, reaches a maximum at a higher temperature, and follows the high-temperature Curie law (13), (14).

The coordinates of the maximum of transverse susceptibility (A.6) are

$$\frac{k_BT_{max}}{|J|} = 0.418778\ldots, \quad \frac{|J|\chi_{\perp,max}^{(1D)}}{N_Ag_\perp^2\mu_B^2} = 0.299919\ldots. \quad \text{(A.12)}$$

For the double-chain Ising model with $z'|J'/J| = 0.1$ (curve *2* in Fig. 2), the transverse-susceptibility maximum has the coordinates (0.480876, 0.288263).

REFERENCES

1. L. J. de Jongh and A. R. Miedema, Adv. Phys. **23**, 1 (1974); **50**, 947 (2001).
2. M. Steiner, J. Villain, and C. G. Windsor, Adv. Phys. **25**, 87 (1976).
3. A. J. Guttmann and I. G. Enting, Phys. Rev. Lett. **76**, 344 (1996).
4. W. P. Orrick, B. G. Nickel, A. J. Guttmann, and J. H. H. Perk, Phys. Rev. Lett. **86**, 4120 (2001); J. Stat. Phys. **102**, 795 (2001).
5. B. M. McCoy, cond-mat/0012193.
6. S.-C. Chang and M. Suzuki, Physica A (Amsterdam) **341**, 299 (2004).
7. R. E. Greeney, C. P. Landee, J. H. Zhang, and W. M. Reiff, Phys. Rev. B **39**, 12200 (1989).
8. C. P. Landee, R. Kuentzler, and J. J. M. Williams, J. Appl. Phys. **67**, 5604 (1990).
9. R. S. Rubins, T. D. Black, and A. Sohn, Phys. Rev. B **49**, 15366 (1994).
10. R. S. Rubins, A. Sohn, and T. D. Black, Phys. Rev. B **61**, 11259 (2000).
11. K. Takeda, S. Matsukawa, and T. Haseda, J. Phys. Soc. Jpn. **30**, 1330 (1971).
12. S. Foner and R. B. Frankel, J. Chem. Phys. **68**, 4781 (1978).
13. L. A. Bosch, G. J. P. M. Lauwers, K. Kopinga, *et al.*, J. Phys. C **20**, 609 (1987).
14. M. Elmassalami and R. C. Thiel, Physica B (Amsterdam) **168**, 137 (1991).
15. H. T. Wittveen, W. L. C. Rutten, and J. Reedijk, J. Inorg. Nucl. Chem. **37**, 913 (1975).
16. H. J. M. de Groot, R. C. Thiel, and L. J. de Jongh, in *Proceedings of International Workshop on Magnetic Excitations and Fluctuations, San Miniato, Italy, 1984*, Ed. by S. W. Lovesey, U. Balucani, F. Borsa, and V. Tognetti (Springer, Berlin, 1984), p. 21.
17. H. J. M. de Groot, L. J. de Jongh, R. C. Thiel, and J. Reedijk, Phys. Rev. B **30**, 4041 (1984).
18. Y. Tanaka and N. Uryû, Phys. Rev. B **21**, 1994 (1980).
19. P. Butera and M. Comi, Phys. Rev. B **65**, 144431 (2002).
20. M. Campostrini, A. Pelissetto, P. Rossi, and E. Vicari, Phys. Rev. E **65**, 066127 (2002).
21. H. Arisue and T. Fujiwara, Nucl. Phys. B (Proc. Suppl.) **119**, 855 (2003).
22. F. Harbus and H. E. Stanley, Phys. Rev. B **7**, 365 (1973).
23. D. Hansel, J. M. Maillard, J. Oitmaa, and M. J. Velgakis, J. Stat. Phys. **48**, 69 (1987).
24. R. Navarro and L. J. de Jongh, Physica B (Amsterdam) **94**, 67 (1978).
25. M. E. Fisher, in *Proceedings of the 1970 E. Fermi International School on Physics of Critical Phenomena*, Ed. by M. S. Green (Academic, New York, 1971), Vol. 51, p. 1.
26. M. N. Barber, in *Phase Transitions and Critical Phenomena*, Ed. by C. Domb and J. L. Lebowitz (Academic, London, 1983), Vol. 8, p. 145.
27. V. Privman, in *Finite Size Scaling and Numerical Simulation of Statistical Systems*, Ed. by V. Privman (World Sci., Singapore, 1990), p. 1.
28. R. J. Kubo, J. Phys. Soc. Jpn. **12**, 570 (1957); R. Kubo, in *Problems of Quantum Theory of Irreversible Processes*, Ed. by V. L. Bonch-Bruevich (Inostrannaya Literatura, Moscow, 1961), p. 39.
29. R. Kubo, in *Thermodynamics of Irreversible Processes*, Ed. by D. N. Zubarev (Inostrannaya Literatura, Moscow, 1962), p. 345 [in Russian].
30. M. A. Yurishchev, Fiz. Nizk. Temp. **9**, 851 (1983) [Sov. J. Low Temp. Phys. **9**, 442 (1983)].
31. M. A. Yurishchev, Phys. Rev. B **50**, 13533 (1994).
32. F. R. Gantmakher, *The Theory of Matrices*, 4th ed. (Fizmatgiz, Moscow, 1988; Chelsea, New York, 1959).
33. M. A. Yurishchev, Phys. Rev. E **55**, 3915 (1997).
34. M. A. Yurishchev, Zh. Éksp. Teor. Fiz. **125**, 1349 (2004) [JETP **98**, 1183 (2004)].

35. M. E. Fisher, Physica (Amsterdam) **26**, 618 (1960); **26**, 1028 (1960).
36. M. E. Fisher, J. Math. Phys. **4**, 124 (1963).
37. M. Fisher, *The Nature of Critical Points* (Univ. of Colorado Press, Boulder, Colo., 1965; Mir, Moscow, 1968).
38. J. W. Essam and M. E. Fisher, J. Chem. Phys. **38**, 802 (1963).
39. A. J. Guttmann and I. G. Enting, J. Phys. A **26**, 807 (1993).
40. M. Thomsen, Phys. Rev. B **34**, 4762 (1986).
41. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. (Cambridge Univ. Press, Cambridge, 1992).
42. L. Onsager, Phys. Rev. **65**, 117 (1944).
43. R. E. Greeney, C. P. Landee, W. M. Reiff, and J. H. Zhang, J. Appl. Phys. **64**, 5938 (1988).
44. J. W. Stout and R. C. Chisholm, J. Chem. Phys. **36**, 979 (1962).
45. C. J. Hamer, J. Phys. A **33**, 6683 (2000).
46. M. Inoue and M. Kubo, J. Magn. Res. **4**, 175 (1971).
47. M. A. Yurishchev, Fiz. Nizk. Temp. **4**, 646 (1978) [Sov. J. Low Temp. Phys. **4**, 311 (1978)].
48. M. Evangelisti, M. L. Kahn, J. Bartolomé, *et al.*, Phys. Rev. B **68**, 184405 (2003).
49. S. Katsura, Phys. Rev**. 127**, 1508 (1962).
50. D. C. Mattis, *The Theory of Magnetism*, Vol. 2: *Thermodynamics and Statistical Mechanics* (Springer, Berlin, 1985), Sect. 3.6.
51. M. A. Yurishchev, Fiz. Nizk. Temp. **6**, 638 (1980) [Sov. J. Low Temp. Phys. **6**, 307 (1980)].

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 101, No. 6, 2005, pp. 1091–1097.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 6, 2005, pp. 1243–1251.

ORDER, DISORDER, AND PHASE TRANSITIONS IN CONDENSED SYSTEMS

Transition from the Ferromagnetic State to the Spin Glass State in Ordered $Fe_{0.75-x}Al_{0.25+x}$ Alloys and the Temperature Evolution of the Magnetic Structure of the $Fe_{0.70}Al_{0.30}$ Alloy

N. N. Delyagin, A. L. Erzinkyan, V. P. Parfenova, and I. N. Rozantsev
Skobeltsyn Institute of Nuclear Physics, Moscow State University, Vorob'evy gory, Moscow, 119992 Russia
e-mail: delyagin@srd.sinp.msu.ru
Received March 30, 2005

Abstract—The magnetic structure of ordered alloys $Fe_{0.75-x}Al_{0.25+x}$ (x = 0, 0.025, and 0.05) is studied by Mössbauer ^{57}Fe spectroscopy in a temperature range of 5–295 K. An increase in the Al concentration at T = 5 K induces a transition from collinear ferromagnetism (x = 0) to a magnetic structure of the cluster-spin-glass type (x = 0.05). The unexpectedly strong effect of aluminum on the magnetic structure is explained by the anomalously high probability of formation of frustrated magnetic configurations with a large number of Al atoms in the nearest neighborhood of Fe atoms. This anomaly is associated with the establishment of a short-range order, which is a key factor determining the radical change in the magnetic structure in a narrow range of Al concentration. The "intermediate" phase of the $Fe_{0.70}Al_{0.30}$ alloy (100 K < T < 200 K) is a mixed-type magnetically ordered phase whose magnetic structure is determined by the competition of opposite exchange interactions. The nominally "ferromagnetic" phase of this alloy (T > 200 K) is characterized by strong violation of the long-range ferromagnetic order, which is due to the effect of the antiferromagnetic superexchange interaction. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Ordered $Fe_{0.75-x}Al_{0.25+x}$ alloys with a DO_3-type structure exhibit interesting and unexpected magnetic properties in a narrow range of concentrations x = 0–0.05. The $Fe_{0.75}Al_{0.25}$ alloy is a collinear ferromagnet with a high Curie temperature ($T_C \approx 750$ K); however, a slight increase in the Al concentration induces an abrupt low-temperature transition to the spin glass phase (at x = 0.05). The magnetic phase diagram of the $Fe_{0.70}Al_{0.30}$ alloy includes three magnetic phases with different properties [1–4]. At temperatures above $T \sim$ 200 K (up to $T_C \sim 400$ K), the behavior of the magnetic susceptibility corresponds to the ferromagnetic (FM) type of ordering. A decrease in temperature from $T \approx$ 200 K to $T \approx 100$ K is accompanied by a monotonic decrease in the magnetization. A blurred peak observed in the vicinity of $T_{SG} \approx 100$ K indicates a transition to the spin glass (SG) phase. The FM (T > 200 K) and SG (T < 100 K) phases are separated by a broad region of an "intermediate" phase, whose magnetic origin remains unknown. In accordance with the phase diagram proposed in [1–4], the intermediate phase may be paramagnetic (or superparamagnetic); however, this assumption has not been directly confirmed in experiments at the microscopic level.

Important information on the magnetic properties of the ordered $Fe_{0.70}Al_{0.30}$ alloy has been obtained recently by neutron scattering method. The small-angle neutron scattering method was used in [5] for studying the slow spin dynamics typical of SG systems. The measurements were made at temperatures below 100 K (within the SG phase of the $Fe_{0.70}Al_{0.30}$ alloy). In the model proposed in [5], the SG phase contains FM clusters of various size with a random orientation of magnetic moments, as well as regions with rapidly fluctuating magnetic moments (paramagnetic zones). The inelastic neutron scattering method was applied in [6] for studying spin excitations for the SG and FM phases (at 18 and 294 K). It was found that the spin dynamics in these two phases is anomalous. Strong perturbations of the long-range FM ordering were observed for the nominally ferromagnetic phase. Anomalies in the spin dynamics in the SG phase are explained by the formation of FM-type clusters formed at low temperatures as a result of spin correlations over short distances. The results of these publications indicate a complex magnetic behavior of the ordered $Fe_{0.70}Al_{0.30}$ alloy and necessitate more detailed studies at the microscopic level.

Here, we report on the results of investigation of local spin configurations in ordered $Fe_{0.75-x}Al_{0.25+x}$ alloys (x = 0, 0.025, 0.05) and temperature evolution of the magnetic structure of the $Fe_{0.70}Al_{0.30}$ alloy using Mössbauer ^{57}Fe spectroscopy. Mössbauer spectroscopy is an effective method for studying magnetic systems

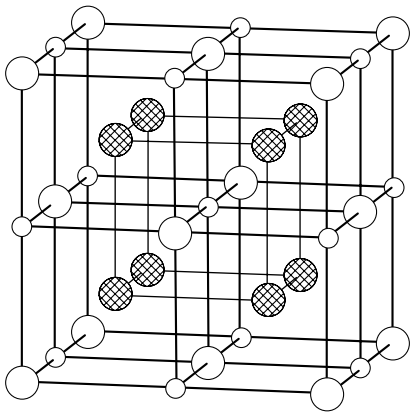


Fig. 1. Crystal structure of the ordered alloy $Fe_{0.75}Al_{0.25}$. Large light circles denote Fe-I atoms, hatched circles denote Fe-II atoms, and small circles are Al atoms.

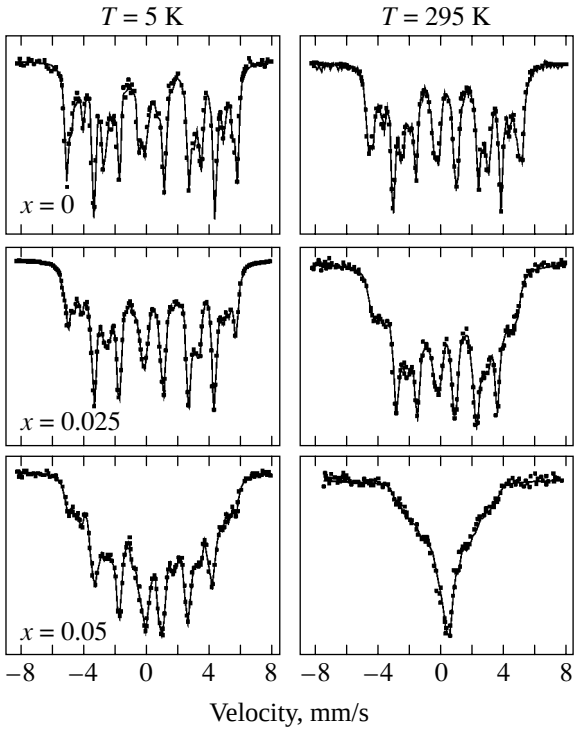


Fig. 2. Mössbauer spectra of the $Fe_{0.75-x}Al_{0.25+x}$ alloys ($x = 0$, 0.025, and 0.05) measured at $T = 5$ K (left) and 295 K (right). Solid curves are the results of approximation of the spectra by the superposition of magnetic subspectra or the hyperfine magnetic field distribution function.

with a complex magnetic structure. The efficiency of the method is determined by the high sensitivity of hyperfine interaction parameters to the properties of local spin configurations. Analysis of hyperfine field distribution (HFD) functions makes it possible to observe various types of local spin configurations and to classify these configurations in accordance with the hyperfine field intensity and temperature dependences of HFD components. In particular, objective information can be obtained on frustrated and paramagnetic states (if such states are present in the system), on the effect of the competition of opposite exchange interactions on the magnetic structure, and on the behavior of magnetic configurations in an external magnetic field.

2. EXPERIMENT

Alloys with different aluminum concentrations ($Fe_{0.75}Al_{0.25}$, $Fe_{0.725}Al_{0.275}$, and $Fe_{0.70}Al_{0.30}$) were prepared by arc melting in argon using metals with a purity not worse than 99.98%. The ingots were homogenized at 1200 K over a week. To obtain alloys in the ordered state, the powders prepared from the ingots were annealed at 400 K during a week, after which the temperature was gradually reduced to 300 K during the next week. The Mössbauer absorption spectra for ^{57}Fe were measured in a temperature range of 5–295 K embracing all the three above-mentioned magnetic phases of the $Fe_{0.70}Al_{0.30}$ alloy. Resonance detectors were used to enhance the resonant absorption effect and the resolving power for detecting Mössbauer radiation. The Mössbauer spectra were analyzed using two different procedures. The HFD functions for all spectra were calculated using the histogram technique [7]. In our calculations, we took into account possible differences in isomer shifts for configurations with different values of the hyperfine field. The method for calculating the HFD functions made it possible to use the procedure of direct minimization of χ^2 functional, which rules out the indeterminacy associated with the smoothing procedure. The spectra with a well-resolved structure were also analyzed using the procedure of approximation of spectra by a superposition of discrete magnetic subspectra.

3. RESULTS

Figure 1 shows the crystal structure of the $Fe_{0.75}Al_{0.25}$ alloy in the case of perfect ordering of the DO_3 type (space group $Fm3m$). The unit cell consists of two mutually penetrating sublattices, one of which contains only Fe atoms (Fe-II sites with coordinates (1/4, 1/4, 1/4), while the other contains Fe atoms (Fe-I sites with coordinates (0, 0, 0)) as well as Al atoms (sites with coordinates (1/2, 0, 0)). In the case of perfect ordering, the occupancy ratio of Fe-I/Fe-II sites is 0.5. Fe-I and Fe-II atoms have eight and four Fe atoms as their nearest neighbors, respectively (configurations Fe(8) and Fe(4)). The magnetic moment of an Fe atom strongly depends on the number of neighboring Fe atoms and is approximately $2.5\mu_B$ for an Fe-I site and approximately $1.50\mu_B$ for an Fe-II site [8]. In accordance with the neutron diffraction data [2], excess Al atoms are predominantly localized in the Fe-I/Al sublattice in the case of a deviation from stoichiometry ($x > 0$).

3.1, Mössbauer Spectra of $Fe_{0.75}Al_{0.25}$, $Fe_{0.725}Al_{0.275}$, and $Fe_{0.70}Al_{0.30}$ Alloys at 5 and 295 K

Figure 2 shows the Mössbauer absorption spectra for three $Fe_{0.75-x}Al_{0.25+x}$ alloys ($x = 0$, 0.025, and 0.05) measured at temperatures of 5 and 295 K. It can be clearly seen that a slight increase in the Al concentration strongly affects the spectral structure. At $T = 5$ K, the discrete structure of the spectra is observed for all three alloys; however, alloys with $x > 0$ exhibit a sharp decrease in the intensity of the high-field components of the spectra. For the $Fe_{0.75}Al_{0.25}$ and $Fe_{0.725}Al_{0.275}$ alloys, the discrete structure of the spectrum is preserved at $T = 295$ K also, while no such structure is observed for the $Fe_{0.70}Al_{0.30}$ alloy at this temperature and the intensity of the low-field components (central part of the spectrum) sharply increases. These results clearly indicate that the substitution of Al atoms for even a small part of Fe sites is accompanied by a strong perturbation of the ferromagnetic structure typical of the $Fe_{0.75}Al_{0.25}$ alloy. The results of approximation of the spectra for the $Fe_{0.75}Al_{0.25}$ and $Fe_{0.725}Al_{0.275}$ alloys by the superposition of magnetic subspectra at $T = 5$ K are given in the table. For the stoichiometric $Fe_{0.75}Al_{0.25}$ alloy, about 96% of the entire intensity are determined by the four subspectra (*A*, *B*, *C*, and *D)*, which corresponds to a high degree of alloy ordering. The most intense subspectra with magnetic hyperfine fields (B_{hf}) of 33.5 and 23.9 T correspond to Fe-I and Fe-II sites with eight and four nearest Fe atoms, respectively (Fe(8) and Fe(4) configurations; see, for example, [9, 10]). The presence of *B* and *C* subspectra indicates a slight violation of the perfect long-range atomic order (it is well known that perfect atomic ordering is never attained in the $Fe_{0.75}Al_{0.25}$ alloy). A small part of Al atoms is localized at the sites of the Fe-II sublattice; a corresponding part of Fe-II atoms move to the Al sites of the Fe-I/Al sublattice. As a result, Fe(7) (subspectrum *B*) and Fe(5) (subspectrum *C*) configurations appear. Low-intensity subspectra *E* and *F* correspond to Fe-II sites in configurations Fe(3) and Fe(2) (the origin of these subspectra will be considered below).

The spectrum of the $Fe_{0.725}Al_{0.275}$ alloy at $T = 5$ K acquires new subspectra, and the intensity of the subspectra corresponding to Fe-II sites with $B_{hf} \approx 18$ and 14 T sharply increases. It can be seen from the table that the intensity of the *F* subspectrum (Fe-II(2) configuration) is practically equal to the intensity of the *E* subspectrum (Fe-II(3) configuration). This result cannot be explained by the assumption on the statistical nature of substitution of Al atoms for Fe sites. Even if we assume (in accordance with the results obtained in [2]) that all excess Al atoms are localized in the Fe-I/Al sublattice, the probability of the emergence of the Fe-II(2) configuration, which is expected from the binomial distribution, turns out to be several times lower than the observed probability. The interpretation of this anomaly is of key importance for determining how the Al concentration affects the magnetic properties of Fe–Al alloys since the formation of low-field configurations directly reflects the frustration of exchange bonds, which accompanies the transition to the state with a magnetic structure of the SG type.

Results of approximation of the Mössbauer spectra of the $Fe_{0.75}Al_{0.25}$ and $Fe_{0.725}Al_{0.275}$ alloys by the superposition of magnetic subspectra at $T = 5$ K; B_{hf} is the hyperfine magnetic field and *I* is the relative intensity of the subspectrum (in data processing, relative intensities are normalized to 100%). The errors in determining B_{hf} and *I* are 0.1 T and 1–2%, respectively

	$x = 0$		$x = 0.025$	
	B_{hf}, T	*I*, %	B_{hf}, T	*I*, %
A	33.5	26	33.3	11
B	31.6	13	31.5	10
C	27.9	14	29.8	4
			27.8	10
D	23.9	43	23.9	37
			22.6	10
E	18.5	3	18.3	10
F	14.0	1	14.0	8

The $Fe_{0.70}Al_{0.30}$ alloy at $T = 5$ K exhibit a sharp increase in the number of the low-field HFD components and in their relative intensity. The total intensity of components with $B_{hf} < 20$ T exceeds 40%; components with $B_{hf} < 10$ T (with a total intensity of about 20%) also appear. It is important to note, however, that the HFD function at $T = 5$ K has no components with zero hyperfine field; the minimal observed value of B_{hf} is approximately 2.5 T. At the same time, components corresponding to the FM type of ordering are observed with an appreciable probability. In particular, the relative intensities of the components corresponding to unperturbed configurations Fe-I(8) and Fe-II(4) in the $Fe_{0.70}Al_{0.30}$ alloy are approximately 5 and 17%, respectively. Such an HFD structure corresponds to a magnetic structure of the cluster spin glass type, in which FM clusters separated by walls with a high concentration of frustrated sites are formed. Our results show that FM clusters constitute at least 40–50% of the alloy volume and that sites at the interfaces are characterized by a broad distribution of exchange fields.

3.2. Temperature Dependence of the Mean Hyperfine Field and Isomer Shift

The temperature dependences of the hyperfine field (both mean value of the field and individual HFD components) for alloys with $x = 0$ and 0.025 do not display any noticeable anomalies and are typical of systems with FM ordering. The temperature dependence of the mean hyperfine field $\langle B_{hf} \rangle$ for the ordered $Fe_{0.70}Al_{0.30}$ alloy is shown in Fig. 3. This dependence exhibits typ-

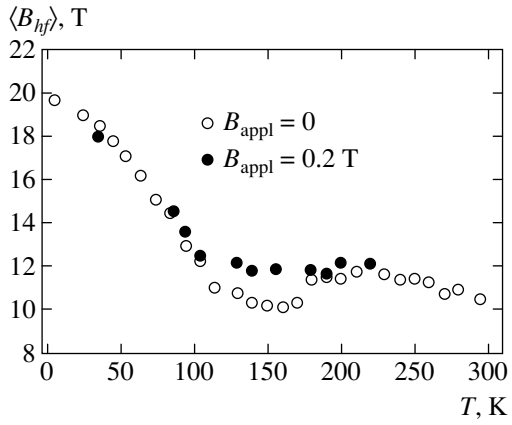


Fig. 3. Temperature dependence of the mean hyperfine magnetic field ⟨B_{hf}⟩ for the $Fe_{0.70}Al_{0.30}$ alloy. Light and dark circles correspond to measurements in zero external magnetic field and in the field B_{appl} = 0.2 T, respectively.

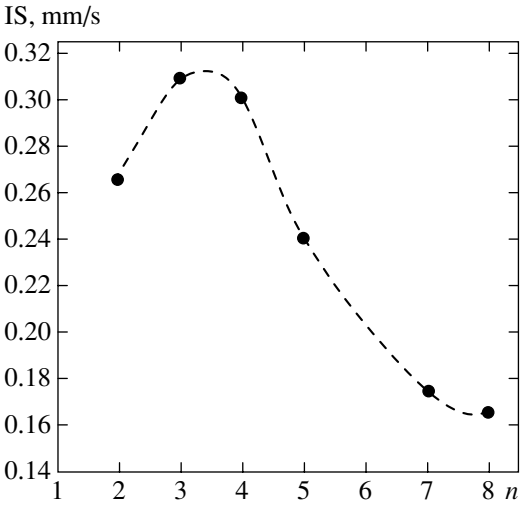


Fig. 4. Dependence of the isomer shift of the Mössbauer line on the number of Fe atoms in the nearest neighborhood.

ical features reflecting the temperature evolution of the magnetic structure of the alloy. In the region of the SG phase ($T < 100$ K), an increase in temperature is accompanied by a rapid decrease in the mean hyperfine field. Upon a further increase in temperature, mean field variations become smoother and do not exceed approximately 2 T in the entire temperature range from 100 to 295 K. In the vicinity of $T \approx 150$ K, a clearly manifested broad minimum of ⟨B_{hf}⟩ is observed. It can be seen from Fig. 3 that this minimum disappears upon the application of a weak magnetic field (B_{appl} = 0.2 T). Beyond this minimum, such a magnetic field does not noticeably affect the value of ⟨B_{hf}⟩.

The behavior of ⟨B_{hf}⟩ in the region of the SG phase resembles the temperature dependence of ⟨B_{hf}⟩ typical of reentrant spin glasses (RSGs) (see, for example, [11]), in which a transition from the SG to FM phase takes place at $T = T_{SG}$. However, the situation for the $Fe_{0.70}Al_{0.30}$ alloy radically differs from that for ordinary RSGs since a transition not to the FM phase, but to a state with a different magnetic structure (intermediate phase) occurs at T_{SG}. In the region of the intermediate phase, the value of ⟨B_{hf}⟩ remains substantial, although considerably smaller than in the case of FM ordering. This phase is obviously neither ferromagnetic nor paramagnetic, but is a "mixed" phase whose magnetic structure is formed as a result of competition between comparable ferromagnetic and antiferromagnetic (AFM) exchange interactions. The broad minimum on the temperature dependence of ⟨B_{hf}⟩ in the vicinity of $T \approx$ 150 K indicates that exchange FM and AFM interactions exhibit different temperature dependences and that the concentration of frustrated sites emerging as a result of competition of the FM and AFM interactions attains its maximal value at $T \approx 150$ K. Since such sites are characterized by strong thermal fluctuations of the magnetic moment, the fraction of Fe atoms with a weak hyperfine field increases. Frustration of Fe sites is accompanied by the formation of a broad distribution of exchange fields. The exchange fields for a certain fraction of Fe sites turn out to be close to zero (which determines the emergence of a low-field peak in the HFD function; see Section 3.3). The Fe atoms occupying these sites are partly aligned in the applied weak magnetic field, which explains the disappearance of the ⟨B_{hf}⟩ minimum in the field of 0.2 T (Fig. 3).

It follows from the temperature dependence of the magnetic susceptibility that the role of the ferromagnetic exchange should increase with temperature. It can be expected that the corresponding change in the balance of the competing interactions should be accompanied by an increase in the value of ⟨B_{hf}⟩. It can be seen from Fig. 3 that a certain increase in the value of ⟨B_{hf}⟩ is observed in the region of transition from the intermediate phase to the nominal FM phase (in the temperature range of 180–220 K); however, the increase in ⟨B_{hf}⟩ is quite small (less than 10%). This means that the AFM component of the exchange interaction in a wide temperature range is not small and does not vanish in the region of the nominal FM phase ($T > 200$ K); for this reason, the transition between the two phases occurs smoothly, without sharp variations in the exchange field distribution. This assumption is confirmed by the results of analysis of the temperature evolution of the HFD (see Section 3.3).

Figure 4 shows the dependence of the isomer shift (IS) of the Mössbauer line on the number n of nearest Fe atoms. (The data were obtained as a result of approximation of the Mössbauer spectra by the superposition of the magnetic subspectra for alloys with x = 0 and 0.025.) It can be seen that the IS rapidly increases upon a decrease in the number of neighboring Fe atoms from 7 to 4. (In particular, this makes it possible to reliably identify the HFD components corresponding to a Fe-I configuration with a large number of neighboring Fe atoms.) Such a behavior of the IS corresponds to the

theoretical model of the Fe–Al atomic interaction, according to which this interaction is local and is determined by the 3*d*(Fe)–3*p*(Al) hybridization [12]. With increasing number of neighboring Al atoms, the screening effect of hybridized electrons on the *s* component of the electron wavefunction becomes stronger, which causes a decrease in the electron density in the region of the nucleus (and hence an increase in the value of the IS).

3.3. Hyperfine Field Distribution Function for the $Fe_{0.70}Al_{0.30}$ Alloy

Figure 5 shows the Mössbauer spectra of the $Fe_{0.70}Al_{0.30}$ alloy at several temperatures in the region of the SG phase and corresponding HFD functions. At temperatures below approximately 50 K, the well-resolved discrete structure of the HFD is preserved. At temperatures above 50 K, the HFD discrete structure is blurred; the intensity of high-field components decreases most rapidly in the region of B_{hf} = 20–35 T. Accordingly, the intensity of the low-field components increases; at T = 85 K, the relative intensity of components with $B_{hf} \le 7$ T is close to 25%. The observed variations of the HFD structure are obviously due to rapid enhancement of thermal fluctuations of the magnetic moment for Fe sites with a suppressed exchange field. It can be seen from Fig. 5 that a local HFD peak formed at T = 105 K in the region of 3–5 T corresponds to sites with the maximal degree of frustration. Suppression of exchange interactions leads to decomposition of the SG cluster structure; however, a considerable fraction of Fe sites remain in states with a well-defined magnetic moment and with values of B_{hf} exceeding 10 T. We can state that the phase emerging at temperatures above the SG transition point is not paramagnetic and contains regions with the FM type of ordering.

The behavior of the HFD function at high temperatures appears unexpected. In accordance with the magnetic phase diagram proposed in [1–4], the $Fe_{0.70}Al_{0.30}$ alloy at T > 180–200 K is ferromagnetic. It would be natural to expect that an increase in temperature would lead to gradual formation of the HFD structure typical of a system with FM-type ordering. (Such a transformation of the HFD can be expected if the exchange AFM interaction decreases much more rapidly than the FM interaction upon an increase in temperature.) Experimental data do not confirm this assumption. In the entire range of temperatures corresponding to the intermediate phase (100–200 K), the HFD function experiences only slight changes associated with variation of the low-field peak height in the region of 3–5 T. Competing exchange interactions exhibit different temperature dependences, but this difference is not as large as expected. Moreover, the HFD function changes insignificantly upon a transition to the nominal "ferromagnetic" phase as well (T > 200 K). The HFD function at temperatures of 150 and 270 K are compared in Fig. 6. It can be seen that the two distributions are quite similar. The main difference is associated with a slight change in the shape of the low-field peak and an increase in the peak height in the region of $B_{hf} \approx 15$ T at T = 270 K. This corresponds to a small change in the value of $\langle B_{hf} \rangle$ upon a transition from the region of the intermediate phase to the nominal ferromagnetic phase (see Fig. 3). The high-field wing of the HFD function (B_{hf} > 10 T) is represented by a broad distribution of

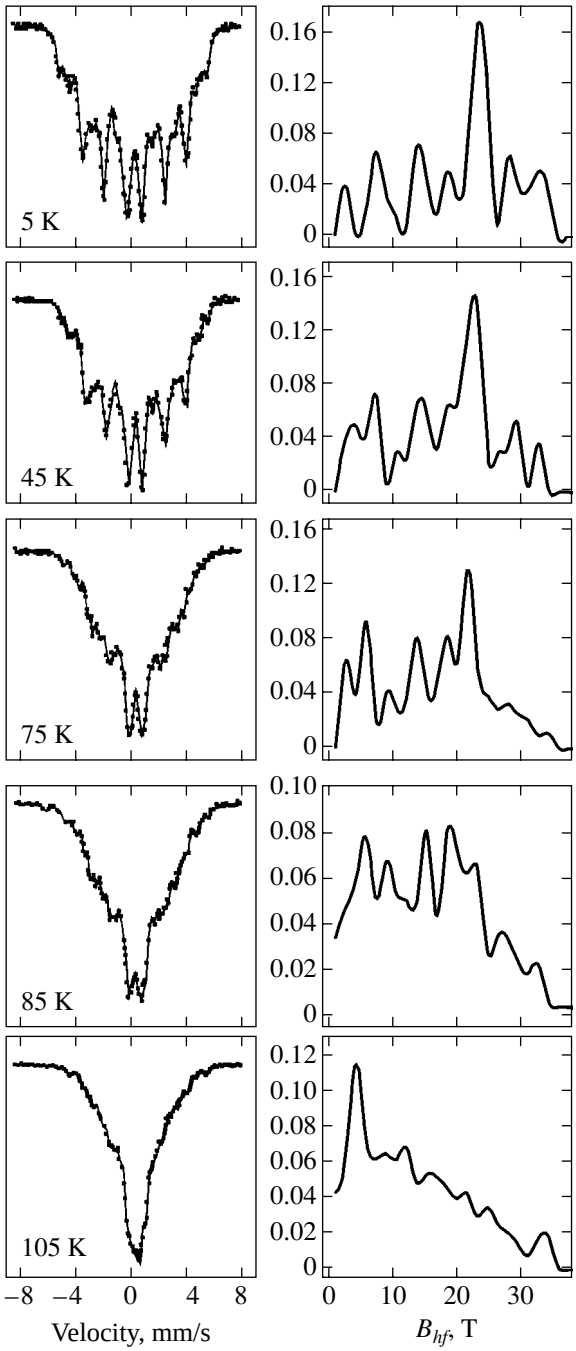


Fig. 5. Mössbauer spectra of the $Fe_{0.70}Al_{0.30}$ alloy at several temperatures in the region of the SG phase (left) and the corresponding hyperfine field distribution functions (right).

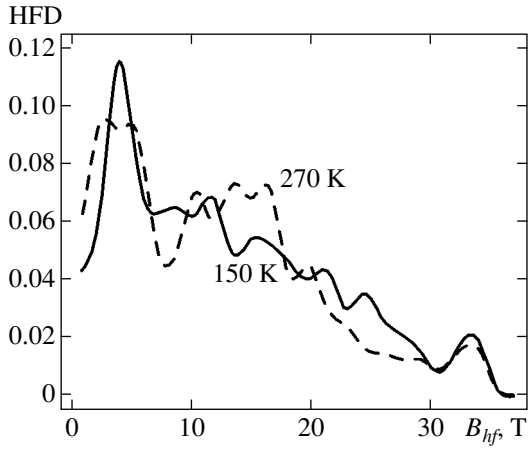


Fig. 6. Comparison of the HFD functions for the $Fe_{0.70}Al_{0.30}$ alloy at $T = 150$ K (solid curve) and $T = 270$ K (dashed curve).

hyperfine fields and has no a well-defined structure that could indicate predominant formation of certain spin configurations of the FM type. The results clearly demonstrate that the role of the AFM exchange is quite important in the entire temperature range studied here. At $T > 200$ K, the ferromagnetic exchange interaction dominates; however, violation of the long-range FM order in the nominal ferromagnetic phase is significant and the magnetic structure of this phase is determined to a considerable extent by the competition of opposite exchange interactions.

4. DISCUSSION AND CONCLUSIONS

4.1. Dependence of the Magnetic Structure of the Alloys at T = 5 K on the Aluminum Concentration

In accordance with the generally accepted assumption, the variations of magnetic properties of Fe–Al alloys are determined by the competition of opposite exchange interactions, viz., the direct FM interaction between neighboring Fe atoms and the indirect (superexchange) AFM interaction via Al atoms (Fe–Al–Fe) [3, 13, 14]. An increase in the Al concentration increases the number of exchange AFM bonds, which must suppress the resultant exchange interaction and gradually transform the FM structure into an AFM structure or a structure of the SG type. This process can be considered, for example, in the framework of the statistical model of distribution of exchange fields, which is based on the percolation theory [15]. However, the change in the magnetic structure occurs in a very narrow concentration range in the alloys studied here, which is difficult to explain in the framework of the statistical approach. In accordance with theoretical estimates (see [3] and literature cited therein), the ratio $-J_{AFM}/J_{FM}$ of competing exchange interaction constants is close to 0.3. It can easily be seen that the necessary condition for the formation of frustrated sites is that six out of eight exchange bonds be antiferromagnetic (Fe(2) configuration). Such a configuration is ruled out for Fe-I sites, but is possible for Fe-II sites (which have four AFM-type bonds for the stoichiometric composition $Fe_{0.75}Al_{0.25}$). However, in the case of the statistical distribution of Al atoms over the sites of the Fe-I/Al sublattice of nonstoichiometric alloys, the probability of formation of the Fe-II(2) configuration is quite low (~0.05 for $x = 0.025$ and ~ 0.15 for $x = 0.05$). These estimates strongly contradict the experimental data, according to which the probabilities of formation of frustrated Fe-II sites ($B_{hf} < 18$ T) is approximately thrice as high (~0.15 for $x = 0.025$ and ~ 0.45 for $x = 0.05$). The anomalously high probability of formation of frustrated Fe-II sites should be regarded as the main reason for the strong influence of Al atoms on the magnetic structure. This anomaly can easily be interpreted from the standpoint of physics. We believe that the only possible explanation is the assumption on the emergence of the short-range atomic order (SRO) resulting in a considerable increase in the probability of formation of configurations for Fe-II sites with a large number of neighboring Al atoms (Fe(2) configurations in the alloy with $x = 0.025$ and the Fe(2) and Fe(1) configurations in the alloy with $x = 0.05$). This mechanism should be treated as a consequence of the strong Fe–Al atomic interaction determined by the formation of hybrid electron orbitals $3d$(Fe)–$3p$(Al) [12]. It should be noted that the tendency towards the establishment of the SRO is observed even in the stoichiometric $Fe_{0.75}Al_{0.25}$ alloy (in the form of a slight violation of regular alteration of Fe and Al atoms in the Fe-I/Al sublattice). In nonstoichiometric alloys with an elevated Al concentration, the SRO becomes the key factor determining the rapid transition from ferromagnetism to a magnetic structure of the spin glass type. Analysis of relative intensities of the HFD components corresponding to the Fe-II(3), Fe-II(2), and Fe-II(1) configurations makes it possible to gain information on the nature of distribution of Al atoms in the establishment of the SRO. To explain the high probability of formation of the Fe-II(2) and Fe-II(1) configurations, we must assume that extended regions with a high local Al concentration are formed in the planes of the Fe-I/Al sublattice (the size of these regions is equal to several lattice constants). The formation of such regions explains the observed intensity ratio I(Fe(2))/I(Fe(3)) (which increases with the effective radius of Al-enriched regions). For a random distribution of excess Al atoms, this ratio must be equal to 0.17 (for $x = 0.025$) and 0.36 (for $x = 0.05$), while the experimentally observed value are approximately equal to 0.9 and 2.0, respectively.

If we assume (in accordance with the result obtained in [2]) that Al atoms are practically absent in the Fe-II sublattice, all exchange bonds for Fe-I sites in the $Fe_{0.70}Al_{0.30}$ alloy remain ferromagnetic (Fe-I–Fe-II). However, for the experimentally obtained high concentration of frustrated Fe-II sites, most Fe-I atoms (about

60%) have at least four frustrated Fe-II atoms with a suppressed exchange field and a small magnetic moment among their nearest neighbors. Attenuation of exchange bonds of the FM type changes the balance of the competing interactions and leads to destruction of the long-range ferromagnetic order in regions with elevated local Al concentration. These regions form the boundaries between clusters, while the FM-type ordering is preserved in regions with a lower local Al concentration. Mössbauer spectroscopy data confirm the results obtained by the neutron scattering method, according to which the low-temperature phase of the $Fe_{0.70}Al_{0.30}$ alloy is a cluster spin glass [2, 4–6]. However, our results do not confirm the model [5], according to which FM clusters at a low temperature are separated by extended paramagnetic zones containing Fe atoms with rapidly fluctuating magnetic moments. Such regions should correspond to states of Fe atoms with zero hyperfine field; however, no such states were observed at low temperatures.

4.2. Temperature Evolution of the Magnetic Structure of the $Fe_{0.70}Al_{0.30}$ Alloy

Analysis of the temperature dependence of the HFD shows that the cluster SG structure of the $Fe_{0.70}Al_{0.30}$ alloy is rapidly destroyed at temperatures above 50 K. This process is primarily determined by the behavior of Fe atoms at the interfaces between the clusters. For Fe sites with a suppressed exchange field, an increase in temperature enhances thermal fluctuations of the magnetic moment (and reduces the observed value of the hyperfine field). This leads to a rapid decrease in the value of $\langle B_{hf} \rangle$ (see Fig. 3) and to an increase in the intensity of low-field HFD components (see Fig. 5).

Our results do not confirm the hypothesis according to which the intermediate phase of the $Fe_{0.70}Al_{0.30}$ alloy (in the temperature range 100–200 K) is paramagnetic. The high relative intensity of components with strong hyperfine fields (close in value to the hyperfine fields in the ferromagnetic phase) clearly indicates the presence of a long-range magnetic ordering. At temperatures above $T_{SG} \approx 100$ K, the relative intensity of HFD components with zero hyperfine field does not exceed 10%, while the total intensity of the low-field components with B_{hf} = 3–5 T does not exceed 30%. It cannot be stated either that the FM clusters determining the properties of the low-temperature SG phase are preserved at $T > 100$ K. Since the HFD exhibits no resolved discrete structure in the region of the intermediate phase, our results do not provide detailed information on the magnetic structure of this phase. At the same time, these data undoubtedly indicate that the intermediate phase is magnetically ordered rather than paramagnetic. We can assume that a network of spins whose mutual orientation is randomly determined by the balance of the exchange FM and AFM bonds at the local level is formed at $T > 100$ K. In all probability, many local spin configurations are noncollinear in this case. This does not follow directly from the experimental data, but makes it possible to explain the absence of a discrete structure in the high-field part of the HFD.

The behavior of the HFD during a transition to the nominal ferromagnetic phase ($T > 200$ K) appears unexpected. The results considered above (see Section 3.3) indicate that this transition is not accompanied by a radical change in the HFD structure and, hence, the magnetic structure of the ferromagnetic phase differs from the structure of the intermediate phase less strongly than expected. Moreover, the expected significant increase in the value of $\langle B_{hf} \rangle$ was not observed either during a transition to the ferromagnetic phase (see Fig. 3). These results lead to the conclusion that violation of the FM order observed earlier by the neutron scattering method is strong and that the magnetic structure of the $Fe_{0.70}Al_{0.30}$ alloy at $T > 200$ K is far from the structure predicted for a collinear ferromagnet. Among other things, this means that although the role of the competing AFM exchange interaction decreases upon an increase in temperature, it remains significant in the entire temperature range studied here.

REFERENCES

1. R. D. Shull, H. Okamoto, and P. A. Beck, Solid State Commun. **20**, 863 (1976).
2. J. W. Cable, L. David, and R. Parra, Phys. Rev. B **16**, 1132 (1977).
3. P. Shukla and M. Wortis, Phys. Rev. B **21**, 159 (1980).
4. K. Motoya, S. M. Shapiro, and Y. Muraoka, Phys. Rev. B **28**, 6183 (1983).
5. K. Motoya, K. Hioki, and J. Suzuki, J. Phys. Soc. Jpn. **72**, 3212 (2003).
6. W. Bao, S. Raymond, S. M. Shapiro, *et al.*, Phys. Rev. Lett. **82**, 4711 (1999).
7. V. P. Parfenova, N. N. Delyagin, A. L. Erzinkyan, and S. I. Reyman, Phys. Status Solidi B **42**, 1465 (2000).
8. D. Satula, L. Dobrzynski, J. Waliszewski, *et al.*, J. Magn. Magn. Mater. **169**, 240 (1997).
9. T. J. Burch, K. Raj, P. Jena, *et al.*, Phys. Rev. B **19**, 2933 (1979).
10. H. Chacham, E. Galvao da Silva, D. Guenzburger, and D. E. Ellis, Phys. Rev. B **35**, 1602 (1987).
11. B. Huck and J. Hesse, J. Magn. Magn. Mater. **78**, 247 (1989).
12. G. P. Das, B. K. Rao, P. Jena, and S. C. Deevi, Phys. Rev. B **66**, 184203 (2002).
13. G. S. Grest, Phys. Rev. B **21**, 165 (1980).
14. B. V. Reddy, S. C. Deevi, F. A. Reuse, and S. N. Khanna, Phys. Rev. B **64**, 132408 (2001).
15. G. Aeppli, S. M. Shapiro, R. J. Birgeneau, and H. S. Chen, Phys. Rev. B **28**, 5160 (1983).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 101, No. 6, 2005, pp. 1098–1105.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 6, 2005, pp. 1252–1259.

ORDER, DISORDER, AND PHASE TRANSITIONS IN CONDENSED SYSTEMS

Spin Reorientation Effects in $GdFe_3(BO_3)_4$ Induced by Applied Field and Temperature¶

S. A. Kharlamova[a], S. G. Ovchinnikov[a], A. D. Balaev[a], M. F. Thomas[b], I. S. Lyubutin[c], and A. G. Gavriliuk[c]

[a]*Kirensky Institute of Physics, Siberian Branch of Russian Academy of Sciences, Krasnoyarsk, 660036 Russia*
[b]*University of Liverpool, L69 3BX, Liverpool, Great Britain*
[c]*Institute of Crystallography, Moscow, 119333 Russia*
e-mail: lyubutin@ns.crys.ras.ru
Received May 16, 2005

Abstract—Magnetic properties of $GdFe_3(BO_3)_4$ single crystals were investigated by ^{57}Fe-Mössbauer spectroscopy and static magnetic measurements. In the ground state, the $GdFe_3(BO_3)_4$ crystal is an easy-axis compensated antiferromagnet, but the easy axis of iron moments does not coincide with the crystal C_3 axis, deviating from it by about 20°. The spontaneous and field-induced spin reorientation effects were observed and studied in detail. The specific directions of iron magnetic moments were determined for different temperatures and applied fields. Large values of the angle between the Fe^{3+} magnetic moments and the C_3 axis in the easy-axis phase and between Fe^{3+} moments and the a_2 axis in the easy-plane phase reveal the tilted antiferromagnetic structure. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The crystal $GdFe_3(BO_3)_4$ belongs to the family of rare-earth borates $RM_3(BO_3)_4$, where R is a rare-earth element and M = Al, Ga, Fe, Sc, and has a trigonal symmetry with the space group $R32$ (D_{3h}^7) [1, 2]. These materials first attracted great interest because of promising nonlinear and laser properties [3–5]. Quite recently, a magnetoelectrical effect was found in $GdFe_3(BO_3)_4$, which suggests that this crystal be considered as a new multiferroic material [6].

The crystal structure of $GdFe_3(BO_3)_4$ can be represented by layers oriented perpendicular to the C_3 axis and consisting of trigonal GdO_6 prisms and smaller FeO_6 octahedrons [2]. The FeO_6 octahedrons are connected by edges and create one-dimensional spiral chains directed along the threefold C_3 axis (Fig. 1). The shortest Fe–Fe interionic distance in chains is about 3.155 Å and that between chains is 4.361 Å, whereas the shortest Fe–Gd distance is 3.746 Å [2]. Exchange interaction between iron ions from different chains is weak and the chains are mutually independent.

Magnetization and magnetic susceptibility measurements have shown that $GdFe_3(BO_3)_4$ is an antiferromagnet with the Néel temperature T_N = 38 K and its magnetic moments are directed along the crystal C_3 axis [7, 8]. It was suggested that magnetic ordering relates to Fe ions, whereas Gd ions are paramagnetic at least down to liquid-helium temperature [7]. However, recent studies of antiferromagnetic resonance [9] indicated that a possible magnetic ordering of the Gd ions at low temperatures can play an essential role in magnetic anisotropy of the crystal and influence the direction of the iron magnetic moment. The competition of the magnetic anisotropy and indirect coupling between Fe–O–Fe chains via Gd^{3+} results in a range of interesting magnetic behavior and, in particular, may be responsible for the spin-reorientation effect observed in [7, 10]. In addition to the low-temperature magnetic transitions, a structural phase transition was observed at 156 K [11]; two electronic and structural phase transitions induced by high pressures at P = 26 GPa and P = 42 GPa and at room temperature were found by optical and X-ray studies [12].

Calculations predicted [6] that the electric polarization and the magnetostriction appearing in $GdFe_3(BO_3)_4$ at low temperatures (the magnetoelectrical effects) are the result of changes in magnetic symmetry during the spin-reorientation transition induced by an applied magnetic field.

Thus, detailed information about distinctive features of the spin-reorientation effect in $GdFe_3(BO_3)_4$ is extremely important for understanding the low-temperature properties and the nature of the magnetoelectrical effect in this material.

In the present paper, in addition to static magnetic measurements, ^{57}Fe-Mössbauer spectroscopy studies were carried out at different temperatures and in an applied magnetic field with single-crystalline samples

¶ The text was submitted by the authors in English.

of $GdFe_3(BO_3)_4$. The temperature and magnetic-field induced spin-reorientation transitions of the Fe^{3+} magnetic moments were found and investigated on both macroscopic and microscopic scales.

2. EXPERIMENTAL

High-quality crystals of $GdFe_3(BO_3)_4$ were grown by the flux method [7]. The crystals were transparent and had green color. The unit cell parameters are a = 9.5491(6) Å and c = 7.5741(5) Å. For the Mössbauer measurements, a platelet with dimensions of 8×5 mm^2 and thickness of 0.3 mm was cut from the bulk single crystal. The C_3 axis was in the plane of the platelet.

Static measurements of magnetization and magnetic susceptibility were carried out using a vibrating-sample magnetometer with a superconducting magnet in the temperature range 4.2–300 K. An external magnetic field up to 7.5 T was applied parallel and perpendicular to the C_3 axis.

The ^{57}Fe-Mössbauer spectra were recorded in transmission geometry with standard spectrometers operating in the constant acceleration regime. Gamma-ray sources of ^{57}Co(Cr) and ^{57}Co(Rh) were used. Three sets of Mössbauer experiments have been carried out with the single-crystalline sample.

In the first set, the Mössbauer spectra were recorded at temperatures of 5, 20, 40, and 300 K in zero applied magnetic field and with the propagation vector $\mathbf{k}_\gamma$ of the Mössbauer gamma rays directed perpendicular to the crystal platelet.

In the second set, the spectra were recorded at 4.2 K in external magnetic fields H = 0, 0.3, and 1.0 T applied in the plane of crystal platelet perpendicular to the direction of the sample C_3 axis. The propagation vector of the Mössbauer gamma rays was directed perpendicular to the crystal platelet; i.e., the C_3 axis, the applied field, and the gamma rays were all mutually perpendicular.

In the third set, the spectra were recorded at 4.2 K in external magnetic fields H = 0, 2.0, and 4.0 T applied in the plane of the crystal platelet, but parallel to the C_3 axis.

3. RESULTS AND DISCUSSIONS

3.1. Summary Results of the Static Magnetic Measurements

Temperature dependences of the direct χ and reciprocal $1/\chi$ magnetic susceptibility with the applied field 0.1 T are shown in Fig. 2. Two anomalies are observed when the field **H** is applied parallel and perpendicular to the crystal C_3 axis. At 38 K, the deviation of $\chi^{-1}(T)$ from the linear law implies a transition from the paramagnetic to an antiferromagnetic state, and the sharp anomaly near 10 K indicates a change of the magnetic structure of $GdFe_3(BO_3)_4$.

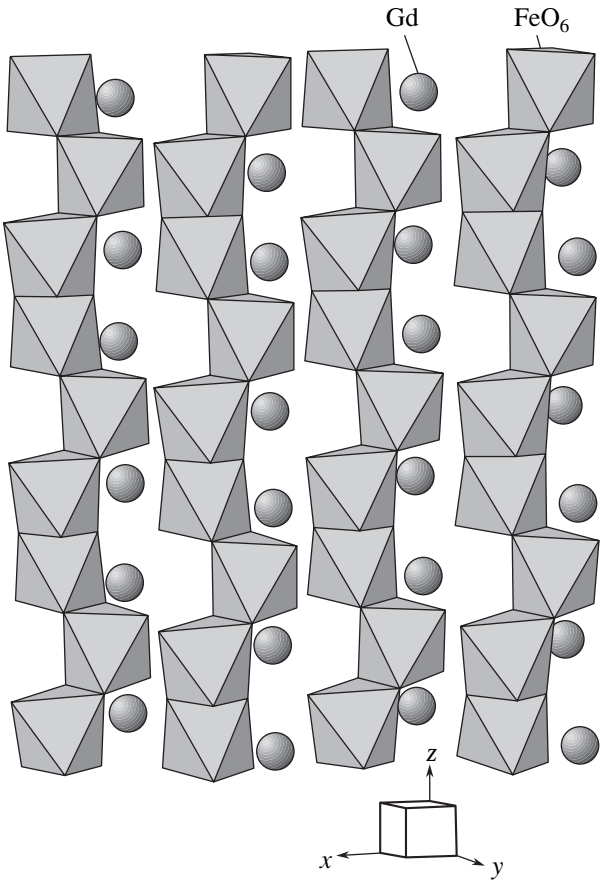


Fig. 1. A fragment of the $GdFe_3(BO_3)_4$ crystal structure showing the oxygen octahedra (the iron ions sites) generating helical chains along the C_3 axis [2].

The field dependences of the magnetization $M(H)$ at 4.2 K are shown in Fig. 3. When the magnetic field is perpendicular to the C_3 axis, the $M(H)$ dependence is almost linear in the range of $0 < H < 3$ T, and magnetization vanishes at zero field (see Fig. 3, inset a). This behavior indicates that the ground state of $GdFe_3(BO_3)_4$ is a compensated antiferromagnet. When **H** is parallel to the C_3 axis, a sharp increase in magnetization was observed at the critical field $H_{reor} \approx 0.6$ T. At this point, M reaches the value typical of that for the case when **H** is perpendicular to the C_3 axis (see Fig. 3). This indicates that a magnetic moment reorientation from the C_3 axis to the plane perpendicular to the C_3 axis occurs. With a further increase in field, an additional anomaly is observed near $H \approx 3.1$ T (see inset b to Fig. 3), which can be attributed to the appearance of the spontaneous magnetic moment induced in the basal plane.

We found that the critical field of reorientation, H_{reor}, decreases with increasing temperature. Figure 4 presents tentative magnetic phase diagrams showing

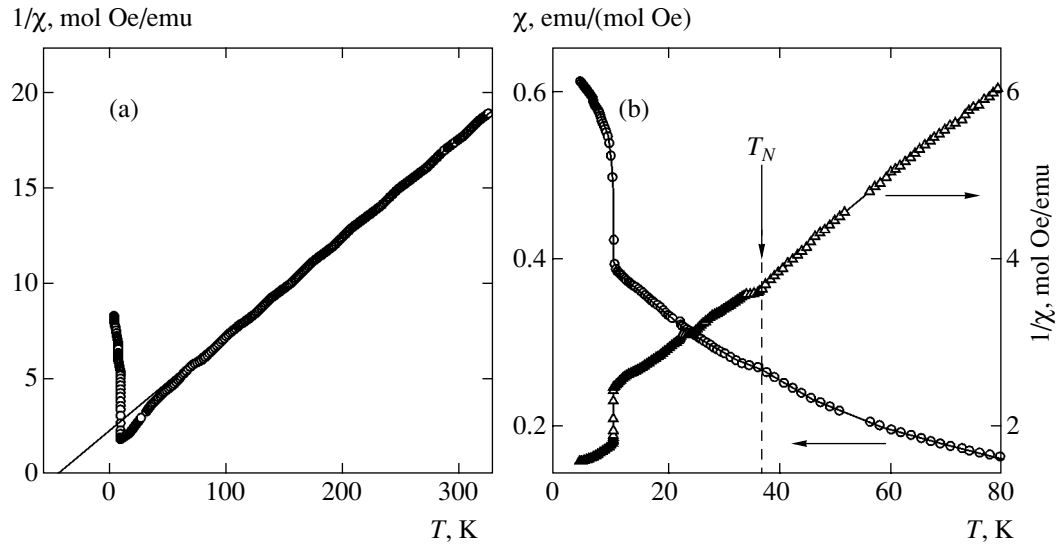


Fig. 2. Temperature dependences of the reciprocal $1/\chi$ and direct χ magnetic susceptibility in the magnetic field $H = 0.1$ T applied parallel (a) and perpendicular (b) to the C_3 axis.

the values of critical fields H_{reor} and corresponding temperatures at which the magnetic moment reorientation occurs when the external field is parallel and perpendicular to the C_3 axis. Along with our data of the static magnetic measurements, the results in [6] from electric polarization and magnetostriction measurements and the results in [9] from the antiferromagnetic resonance measurements are also included in the phase diagram of Fig. 4. One can see that different methods give good agreement for the critical field and its dependence on temperature. The dashed line is a visual guide that separates the two magnetic phases at the reorientation transition from the easy-axis state to the easy-plane state.

3.2. ^{57}Fe-Mössbauer Spectroscopy Results

1. The ^{57}Fe-Mössbauer spectra of the first set of measurements in zero applied field are shown in Fig. 5. At temperatures of 5 and 20 K, the magnetic hyperfine splitting of resonance lines indicates a magnetic ordering of the Fe ions. A single six-line spectrum with a rather narrow linewidth shows that all iron ions occupy equivalent crystal sites, even though the antiferromagnetic ordering implies the existence of at least two iron magnetic sublattices. The magnetic hyperfine field H_{hf} at a ^{57}Fe nuclei and the isomer shift (IS) values are typical of the high-spin Fe^{3+} state (see table). At 40 and 300 K, the spectra show a slightly asymmetric quadrupole doublet characteristic of the iron paramagnetic state. The line broadening at 40 K (see table) is apparently related to a trace of magnetic ordering because the Néel temperature is very close to this temperature. The decrease of the IS value with increasing temperature from 5 to 300 K (see table) is related to the second-order Doppler shift. However, the evident increase of the IS at the transition from the antiferromagnetic to paramagnetic state is an indication of some changes in the chemical bonds and/or the electronic state of the iron ions, which can be related to crystal distortion at the magnetic transition [6].

Special attention should be paid to the behavior of the line intensities in the spectra at 5 and 20 K (see Fig. 5). In general, the line intensities are defined by the probabilities of the Mössbauer transitions between nuclear sublevels and depend on the angle θ between the propagation vector $\mathbf{k}_\gamma$ of the gamma rays and the direction of magnetic hyperfine field H_{hf} at the ^{57}Fe nuclei, which should coincide with the iron magnetic

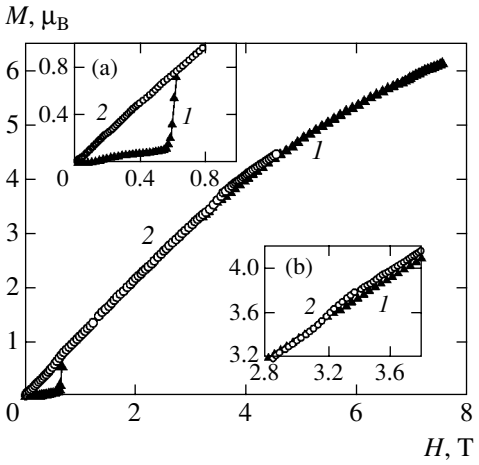


Fig. 3. The field dependences of $GdFe_3(BO_3)_4$ magnetization at 4.2 K in the magnetic field applied parallel (*1*) and perpendicular (*2*) to the C_3 axis. Insets (a) and (b) show the observed anomalies on an enlarged scale.

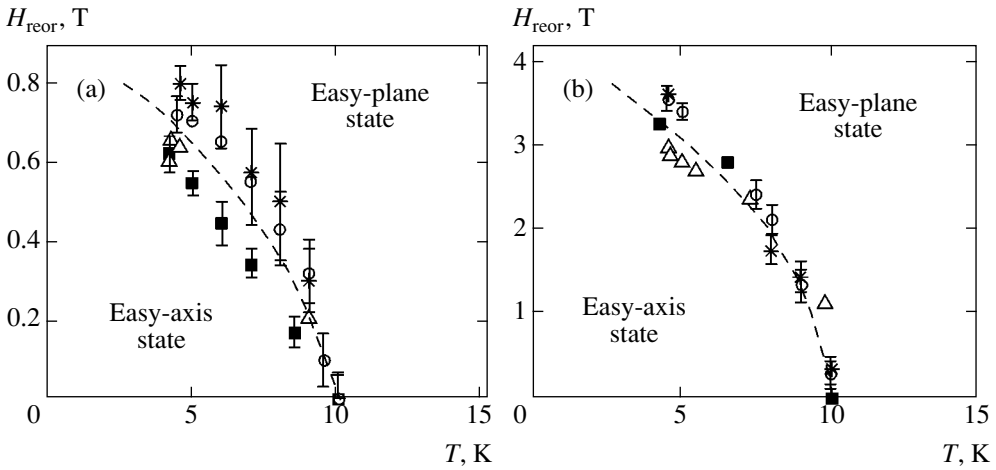


Fig. 4. The tentative magnetic phase diagram for $GdFe_3(BO_3)_4$ showing the values of critical fields H_{reor} and corresponding temperatures at which the iron moments reorientation occurs when the external field is parallel (a) and perpendicular (b) to the C_3 axis. Along with our data (■), the data from the electric polarization (∗) and magnetostriction (○) measurements [6] and the data from the antiferromagnetic resonance measurements (▲) [9] are shown. In both (a) and (b), the dashed line is a visual guide for the eye and separates the two magnetic phases at the reorientation transition from the easy-axis state to the easy-plane state.

moment. For a thin absorber, the intensities I_i (i = 1, 2,…, 6) of a six-line Zeeman spectrum are in a ratio of

$$I_1 : I_2 : I_3 = I_6 : I_5 : I_4 = 3 : \alpha : 1,$$

where

$$\alpha = \frac{4\sin^2\theta}{1 + \cos^2\theta}. \quad (1)$$

In a powder sample, a spherical average of the angle distribution ($\cos^2\theta = 1/3$, $\sin^2\theta = 2/3$) gives

$$I_1 : I_2 : I_3 = I_6 : I_5 : I_4 = 3 : 2 : 1.$$

At a random orientation of a single crystal, the θ value can be obtained from

$$\cos^2\theta = 1 - \frac{2(I_2 + I_1)}{\frac{6}{\sum_{i=1} I_i}}. \quad (2)$$

Thus, it follows from the spectra in Fig. 5 that at temperatures between 5 and 20 K, the iron magnetic moment in $GdFe_3(BO_3)_4$ changes its orientation. From the line intensity ratio in Eq. (2), we found that the iron moments make the angle $\theta = 68 \pm 3°$ with the $\mathbf{k}_\gamma$ vector at 5 K, and this angle changes at 20 K. The low-temperature value of θ indicates that the iron magnetic moments are not in the crystal plane but deviate from it by the angle $\beta = 90° - 68° = 22 \pm 3°$.

Additional information on the direction of the iron magnetic moment can be obtained from the behavior of the quadrupole shift in the Zeeman spectrum below the

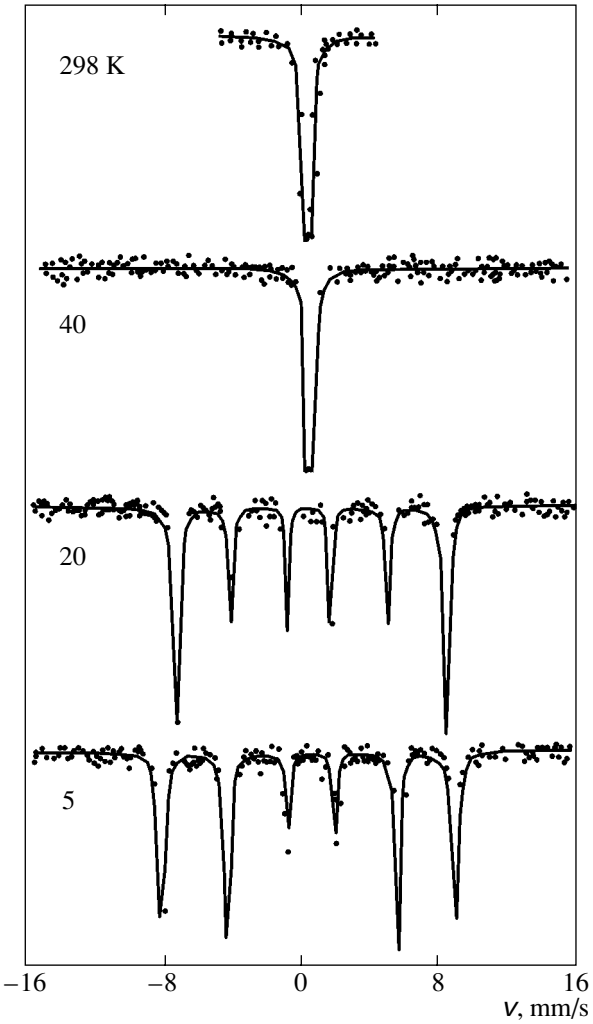


Fig. 5. The ^{57}Fe-Mössbauer spectra of the $GdFe_3(BO_3)_4$ single crystal recorded at temperatures 5, 20, 40, and 298 K in zero applied magnetic field.

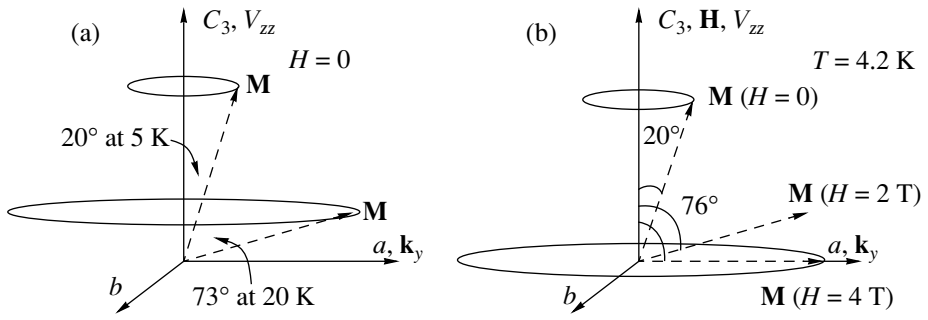


Fig. 6. Effects of reorientation of the Fe magnetic moments in $GdFe_3(BO_3)_4$ deduced from the Mössbauer measurements under the assumption that the hyperfine magnetic field $\mathbf{H}_{hf}$ at the ^{57}Fe nuclei coincides with the iron magnetic moment **M**. Mutual directions of the crystal C_3 axis, the main axis of the electric-field gradient V_{zz}, the applied magnetic field **H**, the gamma-ray propagation vector $\mathbf{k}_\gamma$, and the iron magnetic moments **M** are shown for different temperatures (a) and applied magnetic fields (b). In the case where the antiferromagnetic iron sublattices are collinear, the arrow **M** in (a) and (b) represents the iron antiferromagnetic vector **L**.

Néel temperature. When the magnetic hyperfine interaction is much stronger than the electric quadrupole interaction, the quadrupole shift of the spectral lines observed below the Néel temperature, $(e^2qQ)_{obs}$, and the true quadrupole splitting e^2qQ (which can be obtained above the Néel temperature) are related by

$$(e^2qQ)_{obs} \approx e^2qQ\frac{3\cos^2\varphi - 1}{2}. \qquad (3)$$

Here, Q is the nuclear quadrupole moment and φ is the angle between the direction of the iron magnetic moment and the main axis of electric-field gradient $eq = V_{zz} = \partial^2V/\partial z^2$ [13]. We neglect the asymmetry parameter $\eta = (V_{xx} - V_{yy})/V_{zz}$ in Eq. (3) because of the local C_3 symmetry.

Thus, the angle dependence of $(e^2qQ)_{obs}$ can be used to find the direction of iron moment relative to the crystal axes if the main axis of the electric-field gradient V_{zz} is known. Due to crystallographic reasons, we suppose that in the paramagnetic state of $GdFe_3(BO_3)_4$, the sign of the quadrupole interaction is negative. From the quadrupole shift at $T = 5$ K (Fig. 5), we found the angle φ between Fe moments and V_{zz} direction to be $18 \pm 2°$. It

Hyperfine parameters of the ^{57}Fe-Mössbauer spectra of the $GdFe_3(BO_3)_4$ single crystal: H_{hf} is the hyperfine magnetic field at a ^{57}Fe nuclei, IS is the isomer shift relative to α-Fe at room temperature, QS is the quadrupole shift (splitting), Γ is the linewidth, H is the applied magnetic field, and φ is the angle between the direction of iron magnetic moments and the main axis of electric-field gradient V_{zz}

T, K	H, T	H_{hf}, kOe	IS, mm/s	QS, mm/s	Γ, mm/s	φ
5	0	526.5(2)	0.503(3)	−0.247(6)	0.365(8)	18.3 ± 2.0°
20	0	481.7(2)	0.504(2)	+0.107(4)	0.333(6)	72.8 ± 2.0°
40	0	~0	0.554(2)	0.290(2)	0.431(3)	–
300	0	0	0.390(2)	0.292(2)	0.300(3)	–
External field **H** is perpendicular to the C_3 axis						
4.2	0	527.7(6)	0.485(7)	−0.269(15)	0.39(2)	12 ± 8°
4.2	0.3	527.7(5)	0.497(6)	−0.270(15)	0.37(2)	12 ± 8°
4.2	1.0	527.2(4)	0.497(5)	−0.243(11)	0.38(2)	19.2 ± 2.0°
External field **H** is parallel to the C_3 axis						
4.2	0	527.8(4)	0.510(5)	−0.232(10)	0.380(13)	21.5 ± 2.0°
4.2	2.0	529.8(2)	0.500(4)	+0.120(7)	0.362(10)	76.2 ± 1.5°
4.2	4.0	529.4(2)	0.493(3)	+0.144(6)	0.349(8)	87.3 ± 1.5°

is significant that the angles $\beta = 22 \pm 3°$ and $\varphi = 18 \pm 2°$ agree within their uncertainties and indicate that V_{zz} coincides with the crystal C_3 axis. This suggests that the main axis of the electric-field gradient at an iron nuclei is directed along the local threefold axis of an oxygen octahedral (see Fig. 1). This also shows that at 5 K, the iron moments are not directed precisely along the C_3 axis, but rather at an angle of about 20°.

From the quadrupole shift, we derive an increase of the φ angle to $73 \pm 2°$ at 20 K. If there are no structural transitions between 5 and 20 K, the V_{zz} direction should remain along the C_3 axis. This means that at 20 K, the iron magnetic moments include the angle $73 \pm 2°$ with the C_3 axis (Fig. 6a) and deviate from the basal plane by about $18 \pm 2°$. The intensities of the second and fifth spectral lines also change (Fig. 5), supporting the rotation of the iron spins.

Thus, two independent Mössbauer spectra parameters, the line intensity and the quadrupole shift, indicate that the reorientation of iron magnetic moments occurs in $GdFe_3(BO_3)_4$ at temperatures between 5 and 20 K. In Fig. 6a, mutual directions of the crystal C_3 axis, the main axis of the electric-field gradient V_{zz}, the gamma-ray propagation vector $\mathbf{k}_\gamma$, and the iron magnetic moments **M** are shown for different temperatures under the assumption that the direction of the magnetic hyperfine field $\mathbf{H}_{hf}$ at ^{57}Fe nuclei coincides with the direction of the iron magnetic moments.

2. In the second setup of experiments, the $\mathbf{k}_\gamma$ vector, the C_3 axis, and the external magnetic field **H** were all mutually perpendicular. The spectra of this series recorded at 4.2 K are shown in Fig. 7a. The external fields of 0.3 and 1.0 T applied perpendicular to the crystal C_3 axis do not significantly change the hyperfine parameters at 4.2 K (see table). The iron magnetic sublattices do not lead to the absorption line split and the values of the magnetic hyperfine field H_{hf} are unchanged (see table). This correlates with the suggestion that **H** and $\mathbf{H}_{hf}$ are almost perpendicular. The areas of spectral lines 2 and 5 show that the fields $\mathbf{H}_{hf}$ (and the Fe moments) are not perpendicular to the gamma-ray beam but are at angles of about $\theta = 68 \pm 5°$ (at $H = 0$ and $H = 0.3$ T) and $\theta = 65 \pm 5°$ (at $H = 1.0$ T) to the $\mathbf{k}_\gamma$ vector. The values of the angle φ between the Fe moments and V_{zz} (i.e., the C_3 axis), estimated from the quadrupole shift, are the same for all H and close to the value estimated above at 5 K for zero applied field (i.e., φ is near 20°) within experimental error.

Thus, this series of experiments supports the conclusion derived in Section 3.2.1 that in the low-temperature phase of $GdFe_3(BO_3)_4$ near 4.2 K, the Fe moments deviate from the C_3 axis by an angle of about 20°.

3. The spectra of the third set of experiments with the external field applied along the C_3 axis are shown in Fig. 7b, and the hyperfine parameters are listed in the table. At a fixed temperature of 4.2 K, the changes in intensities of spectrum lines 2 and 5 clearly show the spin reorientation effect induced by the applied field. In the field of 4.0 T, all Fe magnetic moments line up

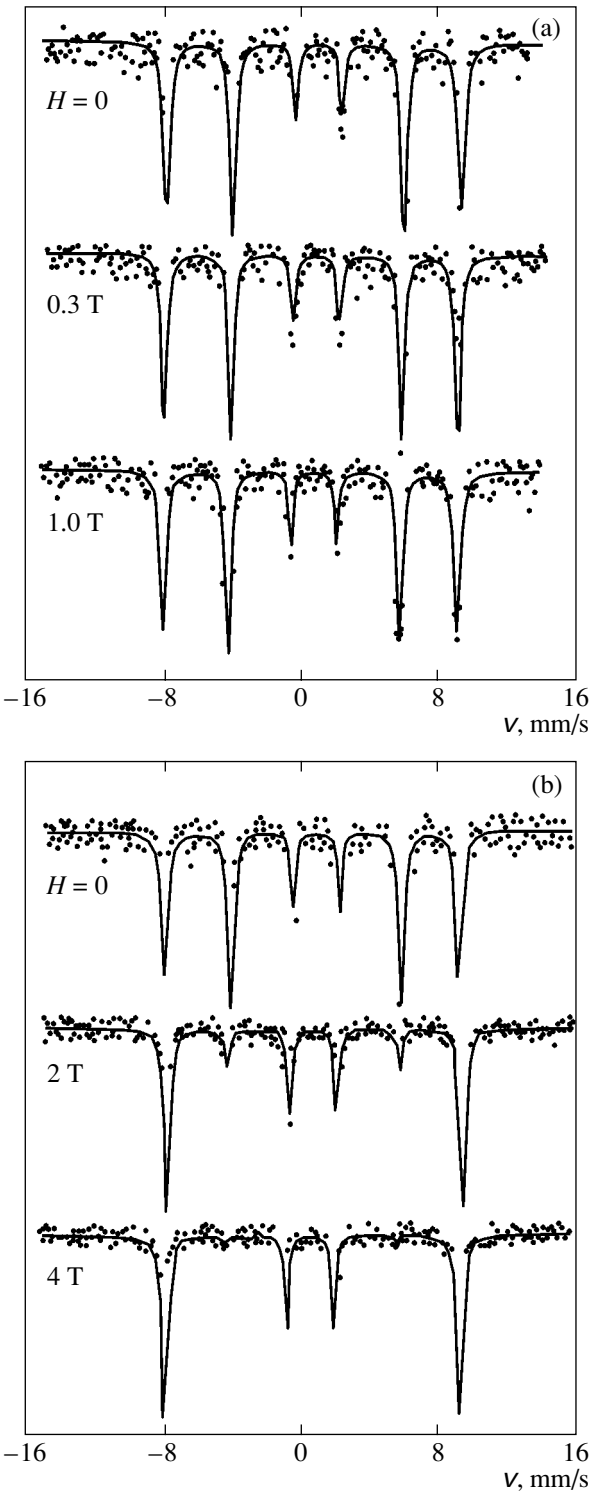


Fig. 7. ^{57}Fe-Mössbauer spectra of the $GdFe_3(BO_3)_4$ single crystal recorded at 4.2 K in external magnetic field applied perpendicular (a) and parallel (b) to the crystal C_3 axis.

along the $\mathbf{k}_\gamma$ vector ($\theta \approx 0$). This means that the angle β between the iron moments and the C_3 axis is near 90°. The behavior of the quadrupole shift supports this observation: the estimated angles φ are about 76.2 ± 1.5° at $H = 2$ T and 87.3 ± 1.5° at $H = 4$ T (see Fig. 6b). The agreement of the β and φ angle values at $H = 4$ T again supports the conclusion made in Section 3.2.1 that V_{zz} direction coincides with the crystal C_3 axis.

Thus, the external field applied along the C_3 axis rotates the iron spins in the (aC_3) plane of the crystal towards the basal plane (see Fig. 6b). At $H = 4$ T, the iron moments are oriented perpendicular to the C_3 axis and also perpendicular to the applied field.

It should be noted that if the antiferromagnetic iron sublattices are collinear, the arrow M in Fig. 6 represents the iron antiferromagnetic vector **L**.

In the basal plane of $GdFe_3(BO_3)_4$, there are three twofold a_2 axes at the angles 120° relative to each other. However, the Mössbauer experiment shows that in spite of the equivalence of the three a_2 axes, the iron moments in the basic plane are aligned completely along only one of the a_2 axes, just the one directed perpendicular to the crystal platelet. It seems that in a thin crystal platelet, the surface anisotropy plays an important role in the iron spin orientation.

In the case of collinear antiferromagnetic ordering of two Fe sublattices, the external field applied along the antiferromagnetic vector would increase ($\mathbf{H}_{tot} = \mathbf{H}_{hf} + \mathbf{H}$) or decrease ($\mathbf{H}_{tot} = \mathbf{H}_{hf} - \mathbf{H}$) the total field H_{tot} at iron nuclei in these sublattices. But the Mössbauer spectra in Fig. 7 show that the applied field does not split the absorption lines and only slightly modifies the H_{tot} value. Rotation of the iron spins normal to **H** with a negligible contribution of the $\mathbf{H}_{hf}$ projection onto **H** explains this behavior.

4. SUMMARY

From the two series of Mössbauer measurements on $GdFe_3(BO_3)_4$ single crystals, we have established that the main axis of the electric-field gradient, V_{zz}, is directed along the crystal C_3 axis.

Our phase diagram in Fig. 4 obtained from the magnetic static measurements shows good agreement with magnetic resonance [9] and electrical polarization and magnetostriction data [6]. The ^{57}Fe-Mössbauer spectroscopy data confirm the spin reorientation transition, first observed in $GdFe_3(BO_3)_4$ by the magnetic static measurements, and give new information on the specific orientation of the iron magnetic moments and the values of angles between the moments and crystal axes at different temperatures and applied fields. In the ground state at 4.2 K, the $GdFe_3(BO_3)_4$ crystal is an easy-axis compensated antiferromagnet. However, the easy axis of the iron moments deviates from the crystal C_3 axis by about 20°.

At about 10 K, the iron magnetic moments reorient spontaneously from the easy axis towards the basal plane. However, at 20 K, the moments are not entirely in the basal plane but deviate from it at about 18°. At 4.2 K, the external field of 1.0 T, applied perpendicular to the C_3 axis, does not influence the iron spins direction. This correlates with the magnetization behavior shown in inset *a* to Fig. 3. The field H applied along the C_3 axis gradually rotates iron moments in the (aC_3) plane toward the basal plane, and at $H = 4$ T, the moments are entirely in the plane. In the basic plane, the iron moments are directed along the crystal a_2 axis, which is perpendicular to the crystal platelet.

Thus, we have found that the magnetic structure of $GdFe_3(BO_3)_4$ is more complicated than it was suggested in [9]. The large values of the angle between the Fe^{3+} magnetic moments and the C_3 axis in the easy-axis phase and between the Fe^{3+} moments and the a_2 axis in the easy-plane phase reveal a tilted antiferromagnetic structure. The origin of this tilting is the competition of the two contributions to the magnetic anisotropy from the Fe^{3+} and Gd^{3+} sublattices. The decreasing of symmetry below the structural phase transition at $T = 156$ K and also below the Néel temperature $T_N = 38$ K [6] provides an additional contribution to the deviation of the magnetic moments from the crystallographic axes.

ACKNOWLEDGMENTS

We are grateful very much to S.N. Sul'yanov for the X-ray analysis of the crystal orientations. We thank A.K. Zvezdin and S.S. Krotov for fruitful discussions. We deeply appreciate the help of W. Sturhahn, who read the manuscript and made valuable comments. This work was supported by the Russian Foundation for Basic Research, project nos. 05-02-16142a, 03-02-16286a, and 04-02-16945a and by the program of the Russian Academy of Sciences "Strongly Correlated Electron Systems."

REFERENCES

1. J. A. Campá, C. Cascales, E. Gutiérrez-Puebla, *et al.*, Chem. Mater. **9**, 237 (1997).
2. N. I. Leonyuk and L. I. Leonyuk, Prog. Cryst. Growth Charact. **31**, 179 (1995); N. I. Leonyuk, Prog. Cryst. Growth Charact. **31**, 279 (1995).
3. D. Jaque, J. Alloys Compd. **323**–**324**, 204 (2001).
4. M. Huang, Y. Chen, X. Chen, *et al.*, Opt. Commun. **208**, 163 (2002).

5. X. Chen, Z. Luo, D. Jaque, *et al.*, J. Phys.: Condens. Matter **13**, 1171 (2001).
6. A. K. Zvezdin, S. S. Krotov, A. M. Kadomtseva, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **81**, 335 (2005) [JETP Lett. **81**, 272 (2005)].
7. A. D. Balaev, L. N. Bezmaternykh, I. A. Gudim, *et al.*, J. Magn. Magn. Mater. C **258**–**259**, 532 (2003).
8. A. D. Balaev, L. N. Bezmaternykh, A. D. Vasil'ev, *et al.*, J. Magn. Magn. Mater. **272**–**276**, E359 (2004).
9. A. I. Pankrats, G. A. Petrakovskiĭ, L. N. Bezmaternykh, and O. A. Bayukov, Zh. Éksp. Teor. Fiz. **126**, 887 (2004) [JETP **99**, 766 (2004)].
10. A. K. Zvezdin, V. M. Matveev, A. A. Mukhin, and A. I. Popkov, *Rare-Earth Ions in the Magnetic Ordered Crystals* (Nauka, Moscow, 1985) [in Russian].
11. R. Z. Levitin, E. N. Popova, R. M. Chtsherbov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **79**, 531 (2004) [JETP Lett. **79**, 423 (2004)].
12. A. G. Gavrilyuk, S. A. Kharlamova, I. S. Lyubutin, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **80**, 482 (2004) [JETP Lett. **80**, 426 (2004)].
13. N. N. Greenwood and T. C. Gibb, *Mössbauer Spectroscopy* (Chapman and Hall, London, 1978).

Journal of Experimental and Theoretical Physics, Vol. 101, No. 6, 2005, pp. 1106–1121.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 6, 2005, pp. 1260–1278.

ORDER, DISORDER, AND PHASE TRANSITIONS IN CONDENSED SYSTEMS

Ground State of Finite Arrays of Magnetic Dots in the Presence of an External Magnetic Field

A. Yu. Galkin[a], B. A. Ivanov[b,c], and A. Yu. Merkulov[c,d]

[a]*Institute of Metal Physics, National Academy of Sciences of Ukraine, Kiev, 03142 Ukraine*
e-mail: a_galkin@i.com.ua
[b]*Institute of Magnetism, National Academy of Sciences of Ukraine, Kiev, 03142 Ukraine*
e-mail: bivanov@i.com.ua
[c]*Taras Shevchenko University, Kiev, 03127 Ukraine*
[d]*FOM-Institute for Plasma Physics Rijnhuizen, Postbus 1207 3430 BE Nieuwegein, The Netherlands*

Received June 6, 2005

Abstract—The ground state of an array of magnetic particles (magnetic dots), which are ordered in a square 2D lattice and whose magnetic moment is perpendicular to the lattice plane, in the presence of an external magnetic field has been analyzed. Such a model is applicable for sufficiently small dots with perpendicular anisotropy that are in a single-domain state and for dots in a strongly inhomogeneous vortex state whose magnetic moment is determined by the vortex core. For the magnetic field perpendicular to the system plane, the entire set of the states has been analyzed from the chessboard antiferromagnetic order of magnetic moments in low fields to the saturated state of the system with the parallel orientations of the magnetic moments of all dots in strong fields. In the presence of the border, the destruction of the chessboard order first occurs at the edges of the system, then near the extended sections of the surface, and finally expands over the entire interior of the array. The critical field at which this simplest state is destroyed is much more weakly than the value characteristic of the ideal infinite system. In contrast to this scenario, the destruction of the saturated state with decreasing field always begins far from the borders. Despite such different behaviors, the magnetic structure in the intermediate range of fields that is obtained with both increasing and decreasing field for finite arrays strongly differs from that characteristic of the ideal infinite system. The role of simple stacking faults of the magnetic dot lattice (such as single vacancies or their clusters) in the remagnetization of the system has been analyzed. The presence of such faults is shown to give rise to the appearance of local destructions of the chessboard antiferromagnetic order at fields that are much weaker than those for an ideal lattice. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION AND FORMULATION OF THE PROBLEM

In recent years, considerable attention in the physics of magnetism has been focused on artificial magnetic materials created using current nanotechnologies. Among these materials are magnetic superstructures containing ferromagnetic elements (films, clusters, and small magnetic particles) that have a characteristic size of tens or hundreds of nanometers and are separated by nonmagnetic layers (see [1–3]). Such materials are important for applications in systems of high-density magnetic recording and can be used to realize logical operations [4]. Moreover, such materials are of interest as essentially new objects of the fundamental physics of magnetism.

One-dimensional superstructures consisting of ferromagnetic metal films that have a thickness of several atomic layers and are separated by layers of nonmagnetic or antiferromagnetic metals are most widely known. Active study of 2D superstructures such as lattices of submicron magnetic particles (which are often called magnetic dots) on a nonmagnetic substrate began in the last decade. Magnetic dots are often manufactured from soft magnetic materials such as iron, nickel, cobalt, and permalloy [1–3, 5, 6], but materials with high anisotropy such as dysprosium [7] are also used. Interaction between single particles in such a system is determined by the magnetic dipole interaction of their magnetic moments; i.e., they constitute a pure realization of dipole magnets, which have been studied theoretically for more than fifty years [8–12]. For dipole-coupled systems, many physical properties that are absent for standard crystal magnets with exchange spin interaction are known, such as the existence of the degenerate ground state with nontrivial degeneration [9–11]. For dipole 2D systems, the Mermin–Wagner theorem is inapplicable and true long-range order is present [13]. Two-dimensional systems with Ising dipoles exhibit a cascade of phase transitions induced by the external magnetic field [14].

Creation and experimental investigation of arrays of magnetic dots introduce a new physical content to theoretical study of dipole coupled systems. For many materials, such as compounds containing rare earth

ions, granulated magnetic materials, diluted solid solutions of paramagnetic ions in nonmagnetic crystals, the situation seems to be similar to that inherent in arrays of magnetic dots. However, arrays of magnetic dots have physical properties that are absent for all of the above systems. First, these systems, in contrast to layered crystals, are truly two-dimensional. For the dipole interaction, in contrast to the case of exchange coupling, the behavior of purely two-dimensional layered systems significantly differs from that of quasi-two-dimensional layered systems [13]. Second, the scale of the dipole interaction between two spins does not exceed several kelvins even for high spins $S = 7/2$, whereas even for the smallest magnetic dots with a volume of 10^5–10^6 nm^3, the characteristic magnetic moment exceeds 10^4 μ_B, where μ_B is the Bohr magneton and the characteristic energy is comparable to or even higher than the energy of thermal motion at room temperature (see the discussion of this problem in [12, 14]). Moreover, for compounds with the high density of rare earth ions, the exchange interaction is weak but is not completely negligible. A quite weak (compared to dipole) ferromagnetic or antiferromagnetic exchange interaction between nearest neighbors leads to different inhomogeneous magnetic states: labyrinth domains [15–22], stripe domains [15–19], or vortices [20–22]. These states significantly differ from the chessboard antiferromagnetic structure that appears due to the dipole interaction [14, 16], and such systems cannot be treated as purely dipole.

A large magnetic moment is also characteristic of granulated magnetic materials, but arrays of magnetic dots, in contrast to the latter materials, are characterized by high spatial regularity. It is worth recalling a new class of materials, namely, molecular crystals of high-spin molecules whose total magnetic moment reaches tens of Bohr magnetons [23]. However, these materials are three-dimensional and it is important that the size of the magnetically active part of a molecule is much smaller than the size of the entire molecule. For this reason, the interaction between magnetic moments is weak.

Thus, arrays of magnetic dots are specific materials with a purely two-dimensional, very regular, lattice structure and long-range dipole interaction between magnetic moments at a sufficiently high temperature [12, 14]. It is reasonable to call them artificial antiferromagnetic materials. Phase transitions induced by the strong magnetic field, including the spin-flop transition known for a crystalline antiferromagnet, are possible in them [24, 25]. However, there is a specific point common for one- and two-dimensional superstructures. All superstructures are large (to hundred layers in the one-dimensional case and tens of thousands dots in the two-dimensional case), but finite systems. The border elements (surface) are expected to play much more considerable role for them in the formation of the properties of the transition. Long ago, Mills [26] and Keffer and Chow [27] found that in antiferromagnets in which the border of a crystal contains spins of only one sublattice, the spin-flop transition begins near the border at a field that is weaker by a factor of $\sqrt{2}$ than the transition field in the interior. It is clear that the possibility of realizing this interesting phenomenon is primarily determined by the quality of the surface and it is difficult to observe this phenomenon for the real (not atomically smooth) surface of the crystal. Probably for this reason, attempts to experimentally find the surface spin-flop transition at the border of a single crystal have been unsuccessful [28]. However, such a transition was observed for one-dimensional superstructures in the form of multilayer Fe/Cr films with antiferromagnetic interaction between layers grown on an anisotropic MgO(110) substrate [29–31]. Two-dimensional systems provide a higher possibility of realizing localized transitions at the border of the surface system, as well as on various irregularities of arrays of dots (defects).

In this work, we analyze both analytically and numerically the ground state of a square array of magnetic dots in the presence of the external magnetic field for both finite arrays of dots and arrays with irregularities. The model is described in Section 2. We primarily consider the Ising case, where the state of a single dot can be described by a certain effective magnetic moment, which is perpendicular to the plane of the system. Such a simplification is applicable for sufficiently small dots with perpendicular anisotropy that are in a single-domain state [6, 7], but it can also be applied for dots in a strongly inhomogeneous vortex state (see [32, 33] and below). For the magnetic field perpendicular to the system plane, the entire set of the states is analyzed from the chessboard antiferromagnetic order of magnetic moments in weak fields to the ferromagnetic state of the system with the parallel orientations of the magnetic moments of different dots in strong fields. For the infinite system, such a model exhibits a rich set of states at intermediate magnetic field strengths [14]. As will be shown in Sections 3 and 4, in the presence of the border or stacking faults in the lattice, the critical fields at which the indicated simple phases lose stability are much weaker than those for the infinite system. In Section 5, intermediate magnetic structures are found using numerical simulation. The forms of the structures in the intermediate field range strongly differ from those characteristic of the ideal infinite system and are considerably determined by the form of the array. The results are discussed in Section 6.

2. DESCRIPTION OF THE MODEL

To describe the system, we suppose that the state of the array is represented by the set of the total magnetic moments $\boldsymbol{\mu}_\mathbf{l}$ of dots. We consider a system of the magnetic moments of dots located at the sites $\mathbf{l} = a(n\mathbf{e}_x + l\mathbf{e}_y)$ of the square lattice, where a is the lattice constant, n and l are integers, and $\mathbf{e}_i$ is the unit vector of the corre-

sponding axis. The Hamiltonian of the system that includes the dipole interaction of such a system of magnetic moments in the presence of both the external magnetic field and magnetic anisotropy for each dot can be written in the form [34]

$$\hat{H} = \sum_{\mathbf{l}\neq\mathbf{l}'} \frac{\boldsymbol{\mu}_\mathbf{l}\cdot\boldsymbol{\mu}_{\mathbf{l}'} - 3(\boldsymbol{\mu}_\mathbf{l}\cdot\boldsymbol{\nu})(\boldsymbol{\mu}_{\mathbf{l}'}\cdot\boldsymbol{\nu})}{|\mathbf{l}-\mathbf{l}'|^3} - \sum_\mathbf{l}[\beta(\boldsymbol{\mu}_\mathbf{l}\cdot\mathbf{e}_z)^2 + \boldsymbol{\mu}_\mathbf{l}\cdot\mathbf{H}]. \quad (1)$$

Here, the first term describes the dipole interaction between the magnetic dots, $\boldsymbol{\nu} = (\mathbf{l}-\mathbf{l}')/|\mathbf{l}-\mathbf{l}'|$, β is the anisotropy constant for a single dot, and the external field $\mathbf{H}$ is assumed to be parallel to the normal vector to the system plane. Anisotropy is supposed to be uniaxial with the easy axis along the z axis, which is perpendicular to the system plane, so that $\beta > 0$ and the state of a single dot is characterized by perpendicular magnetization; in this case, $\boldsymbol{\mu}_\mathbf{l} = \pm\mu_0\mathbf{e}_z$, where μ_0 is the magnetic moment of one dot. This property is inherent in dots elongated along the z axis (for details, see [6, 12]). Another important system for which the model of Ising dipoles is applicable is an array of magnetic dots in the vortex state. Such a state is realized for approximately circular dots made of soft magnetic materials when the radius is larger than a certain critical value $R \geq R_0$; $R_0 <$ 100 nm for permalloy. For quite thin dots with thickness $L \ll R$ (dots with $L \approx 20$–30 nm are really studied), the magnetization $\mathbf{M}$ can be considered as independent of the z coordinate perpendicular to the magnetic dot plane; i.e., $\mathbf{M} = \mathbf{M}(r, \chi)$, where r and χ are the polar coordinates in the dot plane. In this case, the structure of the vortex is similar to the structure well known for magnetic vortices in two-dimensional easy-plane ferromagnets (see, e.g., [35]). The vortex corresponds to the magnetization

$$\mathbf{M} = M_s\{\mathbf{e}_z\cos\theta + \sin\theta[\mathbf{e}_x\cos(\chi+\varphi_0) + \mathbf{e}_y\sin(\chi+\varphi_0)]\},$$

where M_s is the saturation magnetization and $\theta = \theta(r)$ is the polar angle of magnetization. For magnetic dots, the same distribution is realized with $\varphi_0 = \pm\pi/2$, which gives $\mathrm{div}\mathbf{M} = 0$ and leads to the closure of the magnetic flux, i.e., to a decrease in the magnetostatic energy [32]. Taking $\varphi_0 = \pi/2$ for definiteness, we write

$$\mathbf{M} = M_s[\mathbf{e}_z\cos\theta + (-\mathbf{e}_x\sin\chi + \mathbf{e}_y\cos\chi)\sin\theta]. \quad (2)$$

Two values $\theta(r) = 0$ and π are possible and, correspondingly, $\cos\theta(r=0) = p = \pm 1$ at the center of the dot (at $r = 0$). For $r \gg \Delta_0$, the function $\theta(r)$ exponentially approaches $\pi/2$. Here, $\Delta_0 = \sqrt{A/4\pi M_s^2}$ is the exchange length, where A is the inhomogeneous exchange constant. For permalloy, $\Delta_0 \approx 5$ nm. The discrete index p is called the vortex polarization and has the sense of the π_2 topological charge of the vortex [35]. Thus, out-of-plane magnetization is nonzero only in the core region $r \leq \Delta_0$, and the magnetic moment $\boldsymbol{\mu}$ of the dot may assume two values

$$\boldsymbol{\mu} = p\mu_0\mathbf{e}_z, \quad p = \pm 1, \quad \mu_0 = 2\pi\xi L\Delta_0^2 M_s, \quad (3)$$

where $\xi = 1.361$. Since $\Delta_0 \ll R$, the magnetic moment of the core is much smaller than the value for the saturated dot. As a result, the state of the dot is quadruple degenerate, i.e., double degenerate in $\varphi_0 = \pm\pi/2$ and double degenerate in the core polarization $p = \pm 1$. Taking into account only the dipole interaction between dots in the framework of the model given by Eq. (1), one can consider that all dots are in two states, up and down. Here, the difference between the energies of a magnetic dot in the vortex state and a dot magnetized in the system plane serves as magnetic anisotropy. Detailed experimental and theoretical investigations of the ground state of arrays of dots show that this anisotropy is large enough to ensure the Ising orientation of the magnetic moments of dots (see, e.g., [6, 12] and references therein).

Thus, both above cases, though exhibiting significantly different physical pictures, can be described using model (1) under the assumption that the moments of all dots are parallel or antiparallel to the normal vector $\mathbf{e}_z$ to the system plane. We assume that the external field $\mathbf{H}$ is also parallel to the z axis and perform analytical calculations only for this Ising case. In this case, the sign of the magnetic moment at a given site is determined only by the sign of the projection of the effective field on the z axis and a quite complete analysis can be performed analytically. The numerical analysis is performed without the restriction $\boldsymbol{\mu}_\mathbf{l} \parallel \mathbf{e}_z$ using the total Hamiltonian given by Eq. (1) (see Section 5).

3. EFFECTS OF THE FINITENESS OF THE ARRAY UNDER MAGNETIZATION

For the description of the ground state and magnetization processes for real arrays of magnetic dots (in contrast to the idealized model of the infinite array), the problem of the effect of system borders is most substantial, because in principle it cannot be removed. An important role is played by the array edge, where the effective field of the dipole interaction is expected to be minimal and the corresponding magnetic dots are the most sensitive to the external magnetic field. We restrict our analysis to the simplest case, where the array of dots is square. In this and following sections, we also assume that the array is large enough to ignore the direct effects of the size of the system. We primarily consider only the simplest geometry of the system, supposing that the array has the rectangular shape with the borders parallel to certain translation vectors of the ini-

tial array of magnetic dots. We will consider the (1, 0) and (1, 1) vectors. In this case, analysis of the stability of the most significant states, saturated and antiferromagnetic, can be performed analytically using only one dipole sum, which should be determined numerically, for each of these states.

For the infinite array of dots, analysis of possible states in the presence of the magnetic field is simplified due to the invariance of the system under translation by the translation vector of the array of dots. States periodic with a translation period multiple to the period a of the array of dots were considered in [14]. Among these states, only two states—saturated, which we call ferromagnetic (FM), and chessboard antiferromagnetic (CAFM) stable at weak fields—had the same square symmetry as the initial array of dots. The remaining states realized in the intermediate field range were characterized by lower oblique symmetry. The existence of any translational invariance significantly simplifies the problem and enables one to reduce it to the search for the finite configuration of dots with the minimum energy for given average magnetization, which is in essence the difference between the numbers of up- and down-magnetized dots. In the case under consideration, there is no translational symmetry, but other properties can be used. Indeed, analyzing the effect of the borders, we primarily analyze the behavior of the dots located at the border, treating the remaining dots of the array as a reservoir governing the behavior of the border dots. The set of the dots of the array generates the dipole interaction field $\mathbf{H}_D$ on the border dots with magnetic moment $\boldsymbol{\mu}_0$. This field is given by the formula

$$\mathbf{H}_D = -\partial \hat{H}_D/\partial \boldsymbol{\mu}_0,$$

where $\hat{H}_D$ is the Hamiltonian of the dipole interaction between magnetic dots, which is specified by the first term in Eq. (1). If the total magnetic field $\mathbf{H}_t = \mathbf{H} + \mathbf{H}_D$ acting on this dot is parallel to its magnetic moment, the state of the dot is energetically favorable and stable. If the vector $\mathbf{H}_t$ becomes antiparallel to the vector $\boldsymbol{\mu}_0$ at a certain value H_c of the external field $\mathbf{H}$, the magnetic moment of this dot is reversed. It is expected that, in the immediate vicinity of H_c, this reversal does not lead to change in the state of magnetic dots forming the reservoir. We verify the applicability of this approach by comparing the results with the data of the numerical analysis. In essence, this approach is equivalent to the energy approach, which is used in [14] to analyze the stability of the ferromagnetic and CAFM states (see below).

We start with the simplest case of the square array of dots with sides parallel to the elementary-translation vectors. We consider the most significant cases of the ferromagnetic array for which $\boldsymbol{\mu}_n = \mu_0 \mathbf{e}_z$ and the CAFM state in which the magnetic moment of the dot in the site specified by the numbers n and l is equal to $(-1)^{n+l+1}$. These states in the infinite array are stable

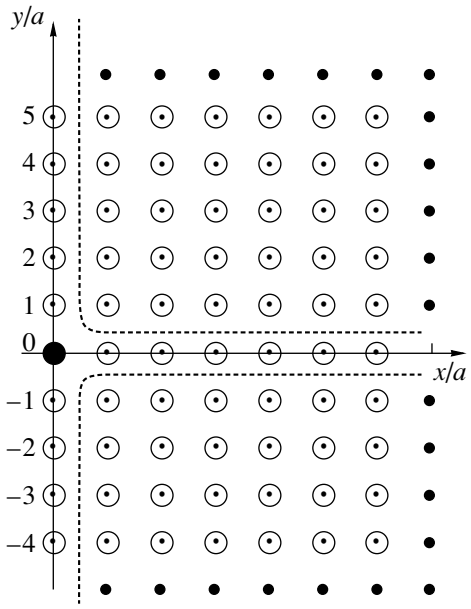


Fig. 1. Semi-infinite lattice of magnetic dots and the choice of the coordinate axes.

for $H > H_1$ and $H < H_0$, respectively, where the characteristic fields H_0 and H_1 are on the order of μ_0/a^3, $H_0 < H_1$ and were calculated in [14]. In what follows, we will sometimes use the dimensionless quantity $h = H(\mu_0/a^3)^{-1}$. In order to calculate the field on a given dot, it is convenient to take the coordinate system with the origin at this dot and with the axes parallel to the border of the array (Fig. 1). The dipole magnetic field at the dot located in the coordinate origin $n = l = 0$ is generated by all other dots of the array and is given by the expression $\mathbf{H}_D = H_D \mathbf{e}_z$, where

$$H_D = -\frac{\mu_0}{a^3}\sum_{n,l}\frac{p_{n,l}}{(n^2+l^2)^{3/2}}. \quad (4)$$

Here, the factor $p_{n,l} = \pm 1$ gives the magnitude of the magnetic moment $\boldsymbol{\mu}_\mathbf{l} = p_{n,l}\mu_0 \mathbf{e}_z$ of the dot with the coordinate $\mathbf{l} = (n\mathbf{e}_x + l\mathbf{e}_y)a$ and n and l are integers at least one of which is not equal to zero.

For all cases of interest, namely, for a magnetic dot located at the edge of the array or at its border and for arrays of dots in the ferromagnetic or CAFM states, a certain common property is easily seen. The sum in Eq. (4) can be expressed in terms of two auxiliary sums: the single sum $\sigma_|$ over dots located at the ray beginning at the coordinate origin (half the coordinate axis) and the double sum $\sigma_\llcorner$ over dots located in one of the array sectors that is bounded by the dashed line in Fig. 1. These sums are defined as

$$\sigma_| = \sum_{n=1}^{\infty}\frac{p_{0,n}}{n^3}, \quad \sigma_\llcorner = \sum_{l=1}^{\infty}\sum_{n=1}^{\infty}\frac{p_{n,l}}{(l^2+n^2)^{3/2}} \quad (5)$$

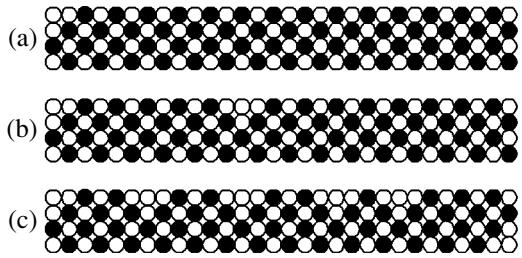


Fig. 2. Magnetic structure of the border and edge of the square array of magnetic dots in the external magnetic field $H = h\mu_0/a^3$, where (a) $h = h^{edge} = 1.562$, at which the reversal of the magnetic moment occurs at the edge of the system; (b) $h = h^{border} = 2.222$, at which the first reversal of the magnetic moment occurs at the border of the system; and (c) $h = 2.264$. The distribution of reversed moments at the border when the field increases is illustrated in panel (c) and Fig. 3. In this and all following figures, magnetic dots magnetized upward and downward are shown by light and dark circles, respectively.

and can be easily calculated. The sum $\sigma_\parallel$ is expressed in terms of the Riemann ζ function, whereas the sum $\sigma_\perp$ can be determined numerically and its values will be presented below. It is easy to see that, for the simplest ferromagnetic and CAFM structures, the field of the dipole interaction at a border dot is determined by these two sums. The field on a magnetic dot located at the infinite rectilinear border (such as the *y* axis in Fig. 1) far from the edges has the form

$$\mathbf{H}^{border} = -\mathbf{e}_z\frac{\mu_0}{a^3}(3\sigma_\parallel + 2\sigma_\perp), \quad (6)$$

and the field on a magnetic dot at the edge of the array is given by the expression

$$\mathbf{H}^{edge} = -\mathbf{e}_z\frac{\mu_0}{a^3}(2\sigma_\parallel + \sigma_\perp). \quad (7)$$

In order to compare this with the results obtained in [14], we write the field on the dot inside the infinite array in a similar form,

$$\mathbf{H}^{vol} = -4\mathbf{e}_z\frac{\mu_0}{a^3}(\sigma_\parallel + \sigma_\perp). \quad (8)$$

We begin with the ferromagnetic state, for which both sums (5) are positive and are equal to

$$\sigma_\parallel^F = \zeta(3) = 1.202057, \quad \sigma_\perp^F = 1.056439,$$

where the latter sum is calculated numerically with the summation up to $n = l = 10^4$. For definiteness, we suppose that all moments are directed upward. In this case, the dipole field is negative on all dots of the array, i.e., $\mathbf{H}_D = -H_D\mathbf{e}_z$, and for the stability of the ferromagnetic state, the presence of an upward external magnetic field exceeding H_D is necessary. It is easily seen that the H_D value at the border points is less than that in the volume. For this reason, when the saturating magnetic field decreases, magnetic dots located at the center of the system are the first dots that lose stability at the field

$$H < H_1 = 4\frac{\mu_0}{a^3}(\sigma_\parallel^F + \sigma_\perp^F).$$

The substitution of σ^F values yields $H_1 = 9.03622\mu_0/a^3$, which was previously obtained in [14]. The characteristic fields at the border or edge are much weaker than H_1:

$$H_1^{border} = 5.718868\frac{\mu_0}{a^3}, \quad H_2^{edge} = 3.4604624\frac{\mu_0}{a^3}, \quad (9)$$

and the ferromagnetic states for them remains stable in a much wider interval of the external magnetic field. This property does not mean that the effect of the border on the formation of the state, where the magnetic moments of some dots are reversed, is negligibly small. As will be shown below, the presence of the border is also substantial in this case. In fact, the border for the ferromagnetic state serves as "repulsive inhomogeneity" for all states containing reversed spins. Numerical analysis confirms this statement (see Section 5). However, a "classical" surface effect, such as the surface spin-flop transition [26, 27, 29–31], where the nucleation of inhomogeneity occurs near the border of the system and then extends to the interior of the sample, is absent for the ferromagnetic state with decreasing external magnetic field.

The above scenario is realized for the CAFM state, which is the ground state in the absence of field and holds stability to a certain critical field. For the infinite system, this critical value is $H_0 = 2.645886\mu_0/a^3$ [14]. Numerical analysis shows that such a CAFM state is also the ground state for finite systems not only of rectangular shape with sides parallel to the (1, 0) and (0, 1) vectors but also of various shapes with acute angles, as well as of circular shape. However, as the field increases, the instability of the CAFM state is developed by reversing the magnetic moments of dots in "weak sections," first at the edge of the array and then at its border (see Fig. 2). In order to demonstrate this behavior, we note that, for the case of the CAFM structure, sums (5) with allowance for the factor $p_{n,l} = (-1)^{n+l+1}$ have different signs:

$$\sigma_\parallel^{AF} = -\frac{3}{4}\zeta(3) = -0.901543, \quad \sigma_\perp^{AF} = 0.2400712.$$

Note that the sum $\sigma_\perp^{AF}$ does not contain any contribution from the nearest neighbors of the edge spin and

is small in absolute value as compared to $\sigma_{\parallel}^{AF}$, $\sigma_{\parallel}^{F}$, and $\sigma_{\perp}^{F}$. This is explained by the faster decrease in the alternating sums corresponding to the CAFM structure. Furthermore, it is clear that the dipole field on a magnetic dot at the edge of the array is parallel to the magnetic moment of the dot, and its magnitude is

$$H_{AF}^{edge} = \frac{\mu_0}{a^3}(2|\sigma_{\parallel}^{AF}| - \sigma_{\perp}^{AF}) = 1.5630142\frac{\mu_0}{a^3}. \quad (10)$$

For a dot with the upward magnetic moment, the external magnetic field, which is supposed to always be directed upward, stabilizes this state. The state of the magnetic dot at the edge of the array in the CAFM state with the downward magnetic moment becomes unstable at the magnetic field $H \approx H_{AF}^{edge}$. Further, the magnetic moment of this dot is reversed at $H \geq H_{AF}^{edge}$. This reversal can be treated as the beginning of the destruction of the CAFM state in the finite array. We note that it occurs at the field that is much weaker (by a factor of 1.7) than the instability field for the infinite system. This state is then stable and does not change until the external field increases to the value

$$H_{AF}^{border} = \frac{\mu_0}{a^3}(3|\sigma_{\parallel}^{AF}| - 2\sigma_{\perp}^{AF}) = 2.2244816\frac{\mu_0}{a^3}, \quad (11)$$

at which the reversals of spins at the border of the array begin. The critical fields presented above are in good agreement with the values obtained in numerical simulation of this system for a quite small size of the system (30×30 dots, see Figs. 2, 3). It is interesting that the field at which the reversal of the magnetic moment in the interior of the system occurs (at this instant, the border is yet incompletely saturated, see Fig. 3) is quite close to the value $h_0 = 2.646$ obtained for the infinite system. All of these facts clearly demonstrate the rapid convergence of dipole sums for antiferromagnetic configurations of the dipoles.

The further evolution of the distribution of magnetic moments proceeds due to the reversal of the moments of dots in the interior of the system (see Section 5). It is worth noting that the border, which contains a large number of moments aligned with the field, quite significantly affects the distribution of the magnetic moments of dots in the interior. Thus, it becomes necessary to analyze other forms of arrays.

As an example, let us discuss a case opposite in some respects to that discussed above. It is an "oblique-square" array, i.e., a rectangular system with borders parallel to the diagonals of the elementary cell of the array of dots, i.e., to (1, 1) vectors (see Fig. 11a). For numerical analysis, we take the array that has a shape close to a square, sides containing only downward and upward magnetic moments, and different configura-

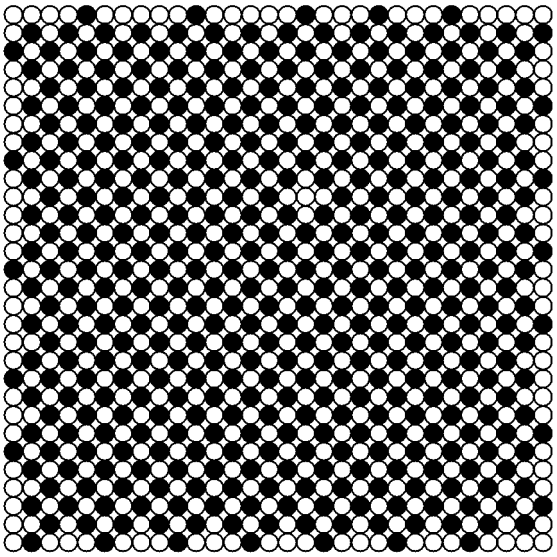


Fig. 3. Magnetic structure of the square array of magnetic dots at the first reversal of the magnetic moment far from the border for the field $h = 2.642$.

tions of magnetic moments at the edges of the system. For the ideal CAFM state, the left and right edges contain dots with the downward and upward moments, respectively, and both these dots belong to the section of the border with the upward and downward moments, respectively. Two remaining edges of the system, upper and lower, contain a compensated pair of magnetic dots. Analysis shows that these two edge states are quite stable against the action of the magnetic field. Figure 4a shows the fragment of such a system in the state with one reversed moment at the edge.

An edge with the downward magnetized magnetic dot and the adjacent border with the same-type dots evidently constitute a weak link of the CAFM structure against remagnetization. The calculation of instability fields of these fragments is similar to the procedure described above. In this case, it is not necessary to numerically calculate two-dimensional dipole sums additional to those already known. Indeed, let us represent the instability field of the edge dot in the form

$$H^{edge} = (2\sigma_{\parallel} + \sigma_{\perp})\mu_0/a^3,$$

where $\sigma_{\parallel}$ is the sum of the fields of the dots located at the border ray and $\sigma_{\perp}$ are the sum of the fields of the remaining dots. In the border ray parallel to the (1, 1) axis, the distance between dots is equal to $\sqrt{2}\,a$ and, therefore, $\sigma_{\parallel} = \zeta(3)/2\sqrt{2} = 0.424991$. It is easy to determine $\sigma_{\perp}$, because $4(\sigma_{\parallel} + \sigma_{\perp})$ is the sum of the fields of all dots in the CAFM structure of the infinite system. This quantity naturally coincides with the above-discussed instability field of the infinite CAFM structure, which is equal to $-4(\sigma_{\parallel} + \sigma_{\perp}) = 2.645886$. From this

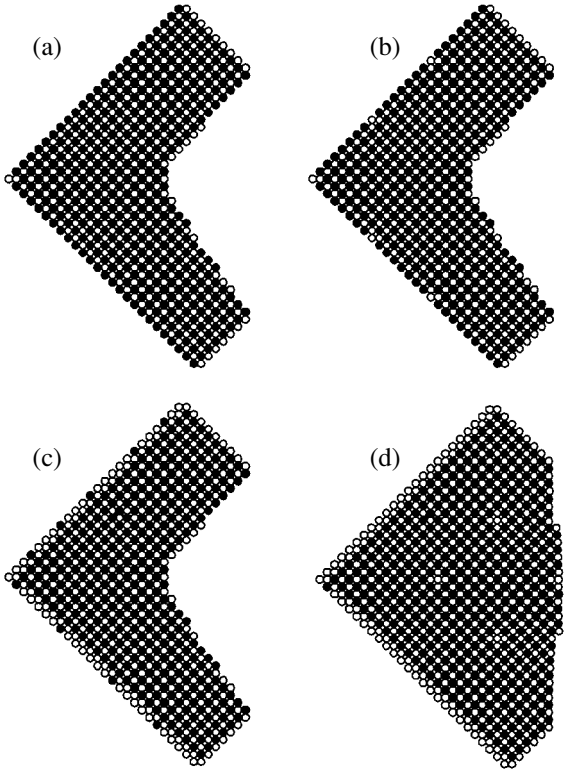


Fig. 4. Evolution of the magnetic structure of the oblique-square array of magnetic dots (see text): (a) $h = h^{edge} = 0.239$, the reversal of the magnetic moment at the edge of the system; (b) $h = h^{border} = 0.899$, the first reversal of the magnetic moment at the border; (c) $h = 2.55$, the end of the magnetization of the border layer; and (d) $h = 2.642$, the first reversal in the interior of the system.

relation, it is easy to find the double sum $\sigma_\perp = -1.0864628$ without numerical calculation. Further, the instability field of the magnetic dot located at the edge, where downward magnetized borders converge, is found as

$$H_\downarrow^{edge} = -(2\sigma_\parallel + \sigma_\perp)\frac{\mu_0}{a^3} = 0.2364802\frac{\mu_0}{a^3}. \qquad (12)$$

As expected, this field is small compared to all of the above-presented instability fields for the array with the border along the (1, 0) and (0, 1) vectors and is one tenth of the value for the infinite system. Correspondingly, the instability field of dots in the downward magnetized border for the array with sides along the (1, 1) vectors is also small:

$$H_\downarrow^{border} = -(3\sigma_\parallel + 2\sigma_\perp)\frac{\mu_0}{a^3} = 0.897952\frac{\mu_0}{a^3} \qquad (13)$$

[cf. Eqs. (12) and (13) with Eqs. (10) and (11)].

Both these values are in good agreement with values obtained numerically (Figs. 4a and 4b). Then, the magnetization of the oblique-square system occurs via almost the same scenario as in the case of the square system with sides along the (1, 0) and (0, 1) vectors. Since the characteristic fields H^{edge} and H^{border} are smaller than the respective values for the direct-square system, the magnetization of the border of this system is realized in a wider field interval. From the value $H_\uparrow^{edge} = 0.2365\mu_0/a^3$ to the volume field $H_0 = 2.642\mu_0/a^3$, only the states of magnetic dots on the initially downward magnetized surface change (Figs. 4b and 4c). The border is almost completely remagnetized to the instant that reversed moments appear in the interior of the system (Fig. 4d).

4. MAGNETIZATION OF THE ARRAY OF DOTS WITH DEFECTS

When analyzing the properties of real arrays of magnetic dots, the question arises about the role of violation of the ideal array structure. Methods used for the production of arrays of magnetic dots enable one to obtain samples with very high spatial regularity. However, for technological applications, it may be necessary to produce arrays with controlled irregularities the simplest of which is a vacancy in the array of magnetic dots. Such vacancies single or united into clusters also evidently constitute a weak link for the destruction of the CAFM order of the array of dots with increasing field, as occurs for the spin-flop transition in low-dimensional antiferromagnets with atomic vacancies [36]. Analytical calculation of the instability field of the CAFM structure near such defects appears to be easy, and the corresponding fields strongly differ from both the volume field H_0 and the above-discussed fields H^{edge} and H^{border}. It will be shown that the magnetization of a given cluster of vacancies occurs via several jumps of the magnetic moment at certain field values. Analyzing such a step curve, one can determine the presence of certain defects in the array of magnetic dots without detailed scanning of the entire array, which can be used for diagnostics of the series of samples.

We start with the analysis of a single vacancy in the CAFM state of the square array. It is evidently sufficient to analyze only one of two equivalent cases, where a dot with the upward or downward moment is removed from the structure. For definiteness, we suppose that the vacancy corresponds to the removal of the dot (or dots) with the upward total moment. Then, dots with the uncompensated downward magnetic moment are located near it. They constitute a weak link of the system in the presence of the positive field, which is discussed here.

Hereinafter, the structure of the vacancy is denoted by the fractional number, where the numerator is the number of removed magnetic dots and the denominator

is their magnetic moment in the CAFM state. For a single vacancy (index 1/1, see Fig. 5), the most "vulnerable" dot is a magnetic dots with a downward magnetic moment that is located immediately near the site from which the magnetic dot is removed. Its magnetic moment is expected to be the first that is reversed. It is evident that the field at this dot is weaker than that for the ideal array by μ_0/a^3, which is the field that is generated at this site by the dot removed from the array. Therefore, the first-instability field is equal to

$$H_{1/1}^{(1)} = H_0 - \frac{\mu_0}{a^3} = 1.645886\frac{\mu_0}{a^3},$$

which is in good agreement with a value of $1.642\mu_0/a^3$ (Fig. 5b). This value is somewhat higher than the value $H^{\text{edge}} = 1.56\mu_0/a^3$ for the direct square array, but is noticeably less than the value for the infinite system. The calculation of the reversal field for the second magnetic moment is also trivial. Formally assuming the existence of an "antidot" with the doubled magnetic moment at the place of the dot with the reversed magnetic moment, we obtain

$$H_{1/1}^{(2)} = H_{1/1}^{(1)} + \frac{2\mu_0}{(2a)^3} = 1.895886\frac{\mu_0}{a^3}$$

(Fig. 5c).

Simple analysis shows that the reversals of subsequent magnetic moments near the single vacancy occur at fields exceeding the volume field H_0. Thus, the existence of the single vacancy gives rise to the appearance of two jumps of the total magnetic moment by a value of $2\mu_0$, which are well separated in the field value. In other words, two minimum "quanta" of the magnetic moment of a given system may be localized at the simplest single vacancy; then vacancy is saturated and is not already involved in the magnetization process.

Let us now consider the compensated double vacancy formed when two neighboring dots that belong to different sublattices in the CAFM structure (index 2/0) are removed. It is intuitively clear that such a defect is weaker than the uncompensated defect. Indeed, elementary calculation shows that the instability field for the magnetic moment is higher than the value for the single vacancy:

$$H_{2/0} = H_0 + \frac{\mu_0}{a^3}\left(1 - \frac{1}{8}\right) = 1.770886\frac{\mu_0}{a^3}$$

(numerical simulation gives a coefficient of 1.753, Fig. 6). An additional reason to treat this defect as weak is that it can localize only one elementary deviation of the magnetic moment. If one moment is reversed, no instabilities near the defect are observed when the field increases from $H_{2/0}$ to H_0.

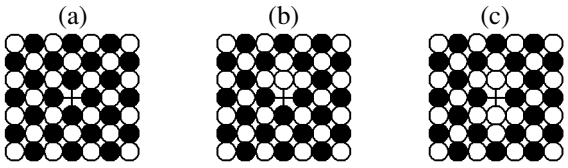


Fig. 5. Evolution of the magnetic structure of the array of magnetic dots near a single vacancy: (a) $h < h_{1/1}^{(1)} = 1.642$, (b) $h = 1.642$, the first reversal, and (c) $h = h_{1/1}^{(2)} = 1.892$, the second reversal.

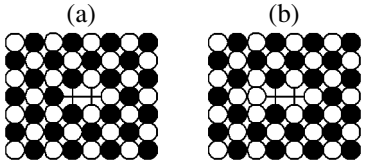


Fig. 6. Magnetic structure near the compensated double vacancy: (a) $h < h_{2/0}^{(1)}$, chessboard order and (b) $h = h_{2/0}^{(1)} = 1.753$, the reversal of the moment.

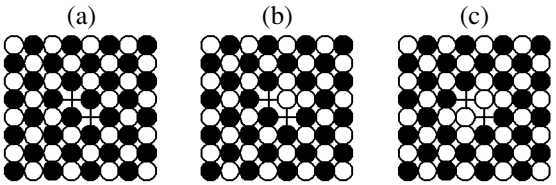


Fig. 7. Evolution of the magnetic structure of the array of magnetic dots near a double uncompensated vacancy: (a) $h < h_{2/2}^{(1)}$, unperturbed state; (b) $h = h_{2/2}^{(1)} = 0.642$, the first reversal; and (c) $h = h_{2/2}^{(2)} = 1.349$, the second reversal.

Double, triple, etc., uncompensated vacancies are stronger and stronger attractive centers. For an uncompensated double vacancy, which is obtained by removing two nearest magnetic dots with moments parallel in the CAFM structure (Fig. 7), the first instability field is already weaker than the field H^{edge} of the reversal of the angular magnetic moment for the most typical square array with the sides along the (1, 0) and (0, 1) axes. In this case, the first reversed magnetic moment is the magnetic moment of the dot equidistant from both empty sites and the field of this reversal is equal to

$$H_{2/2}^{(1)} = H_0 - \frac{2\mu_0}{a^3} = 0.645886\frac{\mu_0}{a^3}.$$

There are two such dots, but the reversal occurs at one of them. After the reversal of the first dot, the field at the

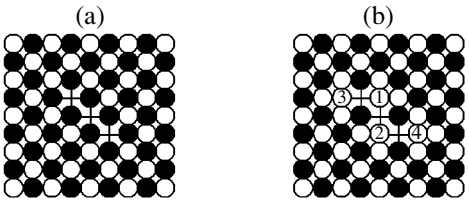


Fig. 8. Evolution of the magnetic structure of the array of magnetic dots near a triple vacancy: (a) the unperturbed state and (b) the state with four reversed moments. Digits 1–4 in the circles indicate the order of reversals of the corresponding moment. According to numerical simulation, these reversals occur at fields $h = h_{3/3}^{(1)} = 0.552$, $h = h_{3/3}^{(2)} = 0.802$, $h = 1.869$, and $h = 1.892$. We point to the closeness of the last two reversals (see text).

second dot increases and its magnetic moment is stable up to the field

$$H_{2/2}^{(2)} = H_{2/2}^{(1)} + 2\left(\frac{\mu_0}{a\sqrt{2}}\right)^3 = 1.352993\frac{\mu_0}{a^3}.$$

The verification of the fields on the remaining dots and numerical simulation show that the role of such a defect ends at this point, see Fig. 7. The double uncompensated vacancy, as well as the single vacancy, may localize two elementary magnetic moments. This defect should be treated as quite strong, because the characteristic fields are noticeably lower than those for the above case of the single vacancy.

The three simple examples discussed above provide sufficient understanding of how to calculate fields at which local reversals of dots near rather complex defects. Let us consider the general problem for a vacancy that is obtained by removing a certain number of magnetic dots from the sites **v** of the ideal lattice. In order to determine the first reversal field, it is sufficient to calculate the field at all dots for which the reversal of the moment is expected. This field at a certain magnetic dot located at the site **n** (**n**th dot), $\mathbf{H}_n = H_n^{(1)}\mathbf{e}_z$, $H_n = h_n^{(1)}\mu_0/a^3$, is determined by the finite sum over vacancies:

$$H_n^{(1)} = H_0 - \frac{\mu_0}{a^3}\sum_{\mathbf{v}}\frac{p_\mathbf{v}}{|\mathbf{v}-\mathbf{n}|^3}, \qquad (14)$$

where the numbers $p_\mathbf{v} = \pm 1$ determine the sign of the magnetic moment in the ideal CAFM structure at the site **v** of the array from which the magnetic dot is removed. The reversal of the magnetic moment in the external field $H^{(1)} = |H_n^{(1)}|$ occurs at the dot for which $H_n^{(1)}$ is negative and its absolute value is minimal. In order to find the next candidate for the reversal, it is necessary to find the fields $\mathbf{H}_n$ on other sites neighboring the vacancy with allowance for the field change at the reversal of the first dot. As a result, for the system with vacancies located at the dots with vectors **v**, where several magnetic moments of the dots located at the sites **l** are already reversed, the field at the **n**th dot is determined by the expression

$$\mathbf{H}_n\{\mathbf{v},\mathbf{l}\} = H_n\mathbf{e}_z, \quad H_n = h_n\mu_0/a^3,$$
$$H_n = H_0 - \frac{\mu_0}{a^3}\sum_{\mathbf{v}}\frac{p_\mathbf{v}}{|\mathbf{v}-\mathbf{n}|^3} + 2\frac{\mu_0}{a^3}\sum_{\mathbf{l}}\frac{1}{|\mathbf{l}-\mathbf{n}|^3}. \qquad (15)$$

Here, we use the fact that a reversal always occurs from the downward position to the upward position and the sign of the corresponding terms is defined. The appearance of a factor of 2 in the sum over the reversed dots was explained above. The reversal of the **n**th dot is possible for $H_n < 0$, which occurs when the external field increases to $|H_n|$.

We now apply this general consideration to describe the system of three uncompensated vacancies located along one line (Fig. 8). In this case, two reversals of magnetic moments occur according to the same scenario as for the double vacancy. They occur for dots equivalent in the CAFM structure, but the moments of these dots are not reversed simultaneously. The distance between them is small enough so that the reversal of the first dot delays the reversal of the second dot. The corresponding fields are given by the expressions

$$H_{3/3}^{(1)} = H_0 - \frac{\mu_0}{a^3}\left(2 + \frac{1}{5}\sqrt{5}\right) = 0.556443\frac{\mu_0}{a^3},$$
$$H_{3/3}^{(2)} = H_{3/3}^{(1)} + \frac{\mu_0}{4a^3} = 0.806443\frac{\mu_0}{a^3}.$$

The next change in the magnetic state of the system has the following feature: two magnetic moments are reversed almost simultaneously at close fields $H_{3/3}^{(3)}$ and $H_{3/3}^{(4)}$:

$$H_{3/3}^{(3)} = H_0 + \frac{\mu_0}{a^3}\left(\frac{3}{4} - \frac{1}{8}\sqrt{2} + \frac{1}{5}\sqrt{5} + \frac{1}{13}\sqrt{13}\right) = 1.873497\frac{\mu_0}{a^3},$$
$$H_{3/3}^{(4)} = H_{3/3}^{(3)} + \frac{2\mu_0}{(a\sqrt{20})^3} = H_{3/3}^{(3)} + 0.0223607\frac{\mu_0}{a^3}.$$

The difference between these fields is caused by the mutual effect of two quite strongly spaced magnetic dots (see Fig. 8). Thus, the triple vacancy localizes four states: two single states and one double state.

The above examples well demonstrate the laws of the appearance of deviations in magnetic moments localized at various defects. For a cluster consisting of

any finite number of vacancies, calculation is reduced to that of a finite sum and, although it becomes rather cumbersome as the number of vacancies increases, it is quite easy.

Let us also present the minimum instability field for the CAFM structure in the presence of extended defects—compensated and uncompensated infinite lines of vacancies. It is clear that, for the uncompensated line of vacancies, in which upward magnetized dots are removed from a certain line parallel to the array vector (1, 1) or (1, –1), the instability field $H^{(1)}_{\infty/\infty}$ is minimal and equal to

$$H^{(1)}_{\infty/\infty} = H_0 - 2\frac{\mu_0}{a^3}\sum_{n=0}^{\infty}\frac{1}{(1+n^2)^{3/2}} = 0.386811\frac{\mu_0}{a^3},$$

which is in good agreement with the value of $0.3895\mu_0/a^3$ obtained in numerical simulation.

For the compensated line of vacancies that is parallel to the (1, 0) or (0, 1) array vector, the instability field of one of the magnetic dots adjacent to the line of vacancies is given by the expression

$$H^{(1)}_{\infty/0} = H_0 - \frac{\mu_0}{a^3}\left[-1 + 2\sum_{n=0}^{\infty}\frac{(-1)^n}{(1+n^2)^{3/2}}\right] = 2.218423\frac{\mu_0}{a^3}.$$

Numerical simulation yields a value of $2.212\mu_0/a^3$.

Finally, for the compensated line of vacancies of the above type that ends with the magnetic dot with the downward moment, the instability of this end occurs at the quite low field

$$H^{\mathrm{end}}_{\infty/0} = H_0 - \frac{\mu_0}{a^3}\sum_{n=0}^{\infty}\frac{(-1)^n}{n^3} = 1.744343\frac{\mu_0}{a^3}.$$

This value is in good agreement with the value of $1.7351\mu_0/a^3$ obtained in numerical simulation.

The results for fields at which the localized violations of the CAFM structure appear at a given defect, are systematized in Fig. 9. For comparison, the horizontal straight lines in this figure show the characteristic fields H_0 of the volume instability, as well as the reversal fields for the magnetic moment at the edge and border of the square array. It appears that two characteristic fields of different natures, namely, the instability field of the magnetic dot near the compensated line of vacancies and the field of the dot reversal at the end of the uncompensated line are close to each other. These values are also close to the border-instability field. For this reason, all these values are shown by the same horizontal dashed line in this figure. The difference between them appears in the third decimal place and this closeness is likely accidental. We emphasize that the characteristic instability fields for the interior and for extended defects (the infinite line of vacancies and the border) are in essence the lower borders of regions where a series of instabilities develops (see, e.g., Figs. 2 and 4). In order to emphasize this fact, the corresponding regions located above these fields ("continuous spectrum of instabilities") are shown by different hatchings.

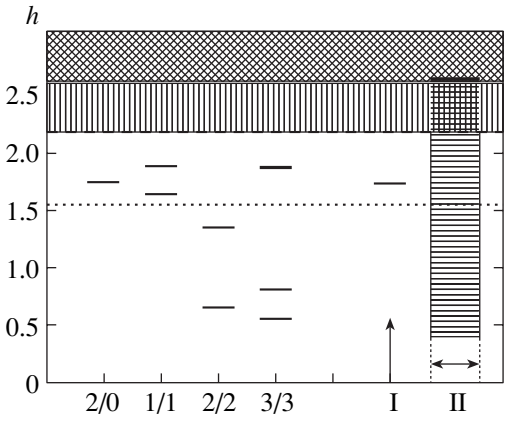


Fig. 9. Instability fields of the CAFM structure of the square array of magnetic dots. The horizontal straight lines are the fields for the defect-free system: the solid line is the bulk-instability field H_0; the dashed and dotted lines are the reversal fields for the magnetic moment at the border and edge of the array, respectively; the short dashes are the characteristic fields for the defects of the array that are shown over the defect index, see text. The data are given for (I) the uncompensated line of vacancies and (II) the end of the compensated line of vacancies. Two close fields for the triple vacancy in the real scale of the figure are represented as the thicker dash. In this scale, the three characteristic fields mentioned in the text are indistinguishable.

5. PROCESSES OF MAGNETIZATION OF THE ARRAY IN INTERMEDIATE FIELDS

In the preceding sections, we showed that the instability field H^{border} of the CAFM structure near the border of magnetic dots is much lower than the value H_0 determined in [14] for the infinite array. In the entire field interval $H^{\mathrm{border}} < H < H_0$, an increase in the magnetic moment occurs due to the reversals of the magnetic moments of dots at the border. Further, for $H \geq H_0$, the volume instability is realized and the presence of the completely (or partially) remagnetized borders of the system almost does not affect this field for the CAFM structure. Analysis of the states of the system for intermediate fields was performed by numerical minimization of total Hamiltonian (1) without the constraint $\boldsymbol{\mu}_n \parallel \mathbf{e}_z$, which was realized by means of an original code.

Numerical analysis was performed with the standard Gauss–Seidel algorithm. The ferromagnetic or

CAFM state realizing the minimum for strong and weak fields, respectively, is taken as the initial state. Calculation was performed beginning with a weak or strong field and with a given step Δh from 2.5×10^{-4} to 10^{-2} and most calculations were performed with steps of $\Delta h = 2.5 \times 10^{-3}$ and 5×10^{-3} (the latter value for larger systems). Simulation corresponds to two regimes usually used in experiment, namely, the magnetization of a sample from the unmagnetized state when the field increases from zero and the demagnetization of the saturated state when the field decreases. Calculation shows quite rapid convergence of the iteration procedure. The code provides automatic calculation of not only energy but also the z projection of the total magnetic moment. The specific feature of this problem is the presence of single-site anisotropy due to which it is necessary to accurately determine the direction of the magnetic moment in a given site at each iteration step with the inclusion of both interaction with neighbors and the effective field of single-site anisotropy. (Without single-site terms, it is sufficient to perform iterations by reversing the magnetic moment along the effective interaction field.)

It is important that the inclusion of single-site anisotropy does not affect the energy of the Ising states found above and that the anisotropy constant β does not appear in all of the above formulas for the critical fields. However, analysis of the general problem with allowance for both the possibility of the deviation of the magnetic moment from the given axis and the presence of single-site anisotropy is fundamentally important for determining the stability of Ising states with respect to the transition to planar states. We are primarily interested in Ising states, which are easily studied analytically, but the inclusion of the transverse components of moments, as well as the oblique field, is not difficult in numerical minimization. Numerical calculations were performed with two anisotropy constants $\beta = 1/a^3$ and $5/a^3$. Most calculations were carried out with $\beta = 1/a^3$, because it is difficult to find a minimum at larger β values. For such a small β value, the transition to planar states is sometimes observed. These states correspond to very complex noncollinear structures, which are considerably determined by the shape of the array. For $\beta = 5/a^3$, the Ising structure is always stable. Thus, the stability of structures obtained in [14] under the assumption $\beta \longrightarrow \infty$ is corroborated at moderate β values.

Our numerical analysis shows that the bulk-instability field for sufficiently small arrays (from 30×30 to 50×50) of various shapes coincides (with an accuracy to three decimal places) with the previously determined value $h = 2.642$. However, the bulk magnetization process in the presence of the field increasing from H_0 is much more sensitive to the existence of borders.

Both investigation of the ideal infinite system and our numerical analysis show that, as the dimensionless field increases by a quite small value, the finite density of magnetic dots with reversed (with respect to the CAFM structure) forms in the array. We characterize this value by the parameter

$$m = M_{\text{tot}}/\mu_0 N, \tag{16}$$

where M_{tot} is the total magnetic moment of the system at a given field and N is the number of the magnetic dots in the system. In the CAFM and saturated ferromagnetic states, $m = 0$ and 1, respectively; i.e., this parameter serves as the reduced magnetization. When a nonzero [even small ($m \leq 0.05$)] density of the reversed dots appears in a sufficiently large system, the effect of the borders on the behavior of the system becomes significant.

Indeed, in an infinite system at a given small m value, the structure is formed due to the dipole interaction of reversed dots against the background of the CAFM structure. For a finite system, there is an additional factor of the mean magnetic field induced by reversed dots. It is expected that the density of reversed dots is approximately constant in the greater part of a sufficiently large system. However, the magnetic field generated by magnetization is not constant over the array space even in this case. As was shown above, this field is maximal at the center, is weaker at the border, and is minimal at the edge of the array, so that these three field values are in a ratio of 9 : 6 : 3 [see Eq. (8)]. Thus, the magnetization-induced "macroscopic" field depends on the distance from the center of the system; moreover, its symmetry presents the geometry of the array.

An additional factor is directly associated with the field of magnetic dots that are located at the border and are upward magnetized for $H \leq H_0$. Already, when the field only slightly exceeds this critical value, the appearance of a superlattice of reversed dots whose density is low for small $H - H_0$ values is expected even in the infinite system. However, such a "network" of reversed dots is in contact in its periphery with the border and is completely sensitive to the presence of the border. For uncompensated borders, which exist in systems such as an oblique square, the difference is stronger and is determined not only by the shape of the array but also by the property of the border, which can be magnetized upward or downward. The effect of the completely or partially magnetized border dictates a certain structure for periphery regions adjacent to it and this structure expands to the central region of the array. Since different sections of the border are oriented differently, such a quasi-constant action of the border sections on the central region is often contradictory and can give rise to frustration effects.

This simple analysis clearly demonstrates the complexity and ambiguity of all factors affecting the magnetic structure of a finite array of dipole-coupled dots for $H > H_0$. In this case, it is reasonable to use direct numerical simulation. We start with the most characteristic example of the square array with sides parallel to

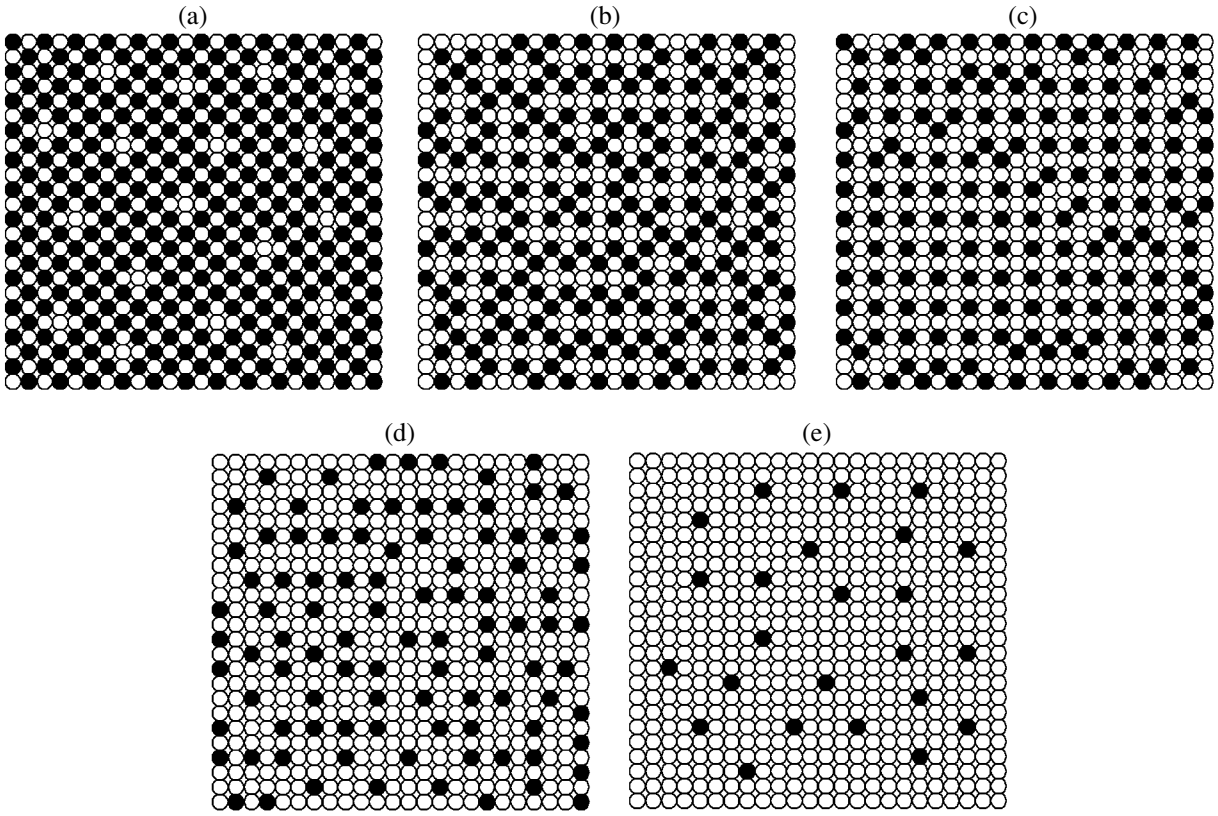


Fig. 10. Magnetic structures of the central region of the square array of magnetic dots for the reduced magnetization m = (a) 1/16, (b) 1/3, (c) 4/9, (d) 2/3, and (e) 11/12 and the dimensionless field h = (a) 2.748, (b) 4.134, (c) 5.154, (d) 7.102, and (e) 8.168.

the (1, 0) and (0, 1) vectors and then briefly discuss other geometries of the system.

Figure 10 shows the central regions of the square array, where surface layers containing the completely magnetized border and adjacent layer are removed for clearness. These data generally corroborate the aforementioned tendency: the structure of the central region of a finite array is formed as compromise due to the competition between several contradictory factors. As a result, the ordered lattice of reversed dots characteristic of the infinite system does not appear even for small m values such as 1/16 (Fig. 10a).

As the field increases further, the number of reversed moments increases and the distance between them becomes much shorter than the size of the system. However, this increase proceeds against the background of the already formed network of deviations from the CAFM structure, which is considerably adapted to the shape of the system and the macroscopic dipole filed generated by magnetization. For this reason, for moderate values m = 0.3–0.5, oblique superlattices, which are characteristic of the infinite system and whose elements are observed for small m values, are not observed. Instead of them, the fragments of the rectangular superlattice of deviations from the CAFM state are formed. The basic element of such a lattice is the line of upward magnetized magnetic dots that is parallel to the (1, 0) or (0, 1) axis (horizontal or vertical sections consisting of open circles in figures). Approximately equal numbers of the horizontal and vertical sections of such lines are clearly seen in a wide m range (see Figs. 10b–10d). For m values close to 1/2, the square lattice covers the greater part of the array (see Figs. 10c and 12a for m = 4/9 and 1/2, respectively). This square lattice should be considered as most favorable with allowance for symmetry dictated by the border. It is worth noting that, for m values close to 1/2 (data for m = 4/9, 1/2, or 5/9 were presented in [14]), various oblique lattices of reversed magnetic moments are realized in the infinite array. Their characteristic feature is the presence of (2, 1) and (1, 2) translation vectors; i.e., they are in poor agreement with closed surfaces with the (0, 1) or (1, 0) borders discussed above, as well as with the (1, 1) and (–1, 1) borders that will be briefly discussed below. As was mentioned in [14], the difference between the energies of different structures with a given m value is small. For this reason, even a small surface effect may strongly change the structure of the lattice and the result obtained above is not surprising.

With a further increase in the field h to 7.5–8, an almost saturated state is formed, where magnetic dots with the downward moment form a structure with quite a low density. In this state, the specific orienting effect

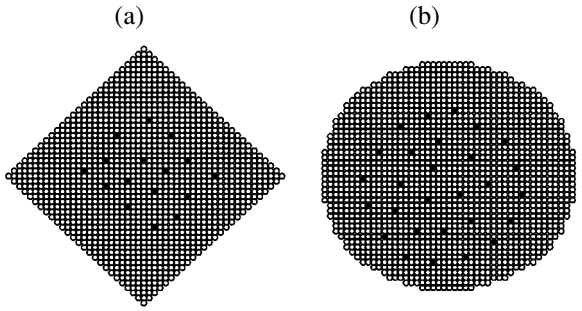


Fig. 11. Magnetic structures arising after the destruction of the saturated state for (a) the oblique square array and (b) approximately circular array.

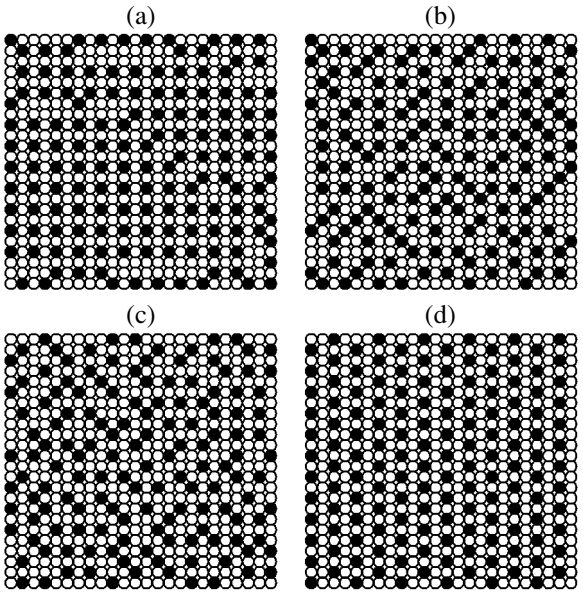


Fig. 12. Magnetic structure corresponding to the reduced magnetization $m = 1/2$ for the central region of the arrays of magnetic dots of the shape of (a) square, (b) oblique square, (c) circle, and (d) structure obtained in [14] for the infinite lattice (shown for comparison).

of the borders is already absent (the borders, as well as the most part of the array, are upward magnetized), but the inhomogeneity of the mean field of the array that is generated by magnetization $m \leq 1$ is completely manifested. In this case, the properties are the same as for $H \geq H_0$: the local short-range order is similar to that arising for lattice states in the infinite system, but more or less amorphous structure really appears for a finite system (see Fig. 10e). The presence of an extremely irregular structure near the saturation field corresponds to experiment (see Fig. 10 in [6]).

The appearance of the same structures accompanies the destruction of the saturated ferromagnetic state, which occurs when the field decreases from H_1. For this process, the effect of the finiteness of the array that is associated with the total inhomogeneous dipole field is clearer. In the saturated state, the field of the magnetic dipole interaction at each dot is negative; i.e., it is directed opposite to the magnetic moment of the dot. For this reason, such a state exists only for finite field values $H > H_1$. The absolute value of the dipole field is maximal at the center of the array and is minimal at its periphery (at the border and particularly at the edges). Therefore, as the field decreases, the magnetic moments of dots at the center of the system are first reversed. However, the nonzero, even very small, density of reversed dots cannot be formed due to their strong dipole interaction. When a dot is reversed, change in the field at neighboring dots is on the order of $2\mu_0/a^3$, which is comparable with the difference between h_1 and h^{border}. Therefore, competition between the interaction of nearest neighbors and the total field generated by all dots of the array is important in the process of the destruction of the ferromagnetic state. Similar to the process of the destruction of the CAFM order in weak fields (but owing to another cause), the border here dictates the structure of the state with a low density of reversed dots for small $H_1 - H$ values. With a further decrease in the field, this structure affects the global symmetry of magnetic states with $m \approx 0.5$. As a result, with both an increase in the field from the CAFM state and a decrease in the field from the ferromagnetic state, similar structures appear in the intermediate region. It is worth noting that the hysteresis effects appear to be weak in such a seemingly nonergodic system. The dependence of the form of the structure realized in intermediate fields on the magnetic field variation step Δh in the simulation process is also negligibly weak in the range from 2.5×10^{-4} to 10^{-2}.

Thus, analysis of the square array shows that the shape of the array strongly affects the distribution of the magnetic moments of dots in the array. To demonstrate that this conclusion is general, we perform numerical simulation for an oblique square array with sides along the (1, 1) and (–1, 1) vectors and for an approximately circular array. Both arrays are cut from the 50×50 lattice. Detailed analysis of numerous numerical data for these two cases is beyond the scope of this paper and we discuss only general properties.

The most characteristic case is the destruction of the saturated ferromagnetic state when the field decreases. As was already mentioned above, the structure arising in this case is a result of the competition between the interaction of nearest dots and interaction of each dot of the array with the weakly inhomogeneous total field of the almost saturated arrays. Analysis of the destruction of the ferromagnetic state for two above-indicated geometries shows that, similar to the direct square considered above, the structure whose general geometry is determined by the shape of the system is formed at the initial stage. The structure of the cloud of reversed dots for H values close to H_1 almost completely reproduces the shape of the array (Fig. 11).

For states with moderate magnetization, the structure is harder, but the shape of the array is again primarily determined by the distribution of the magnetic moments of dots. Figures 12a–12c show the identical square fragments of the arrays of three different shapes described above that are cut from the central part of the array. The field values are chosen such as to ensure the same magnetization $m = 1/2$ in the periphery region after cutting. Figure 12d shows the periodic structure corresponding to the energy minimum of the infinite lattice for $m = 1/2$. The differences between these states are clearly seen. The ideal structure with $m = 1/2$ constitutes a weakly deformed triangular lattice, which is the feature of such structures for many m values [14]. For finite systems, the structure elements in the form of an almost regular hexagon are almost invisible except one such fragment in the right upper quadrant in Fig. 12c for the case of the circular system. From the viewpoint of global properties, the ideal structure has lower symmetry, because it is noninvariant under rotation by 90°. This property is manifested in the orientation of its specific elements, which are the lines that contain upward magnetized magnetic dots and are aligned (Fig. 12d) along the (0, 1) direction. It is clear that the structure rotated by 90° with the same lines along the (1, 0) direction is energetically equivalent to the structure shown in Fig. 12d, which determines the low symmetry of these states. In contrast, such an anisotropy is absent for all structures of real systems. In view of this circumstance, the case of the square array (Fig. 12a) is instructive, where, as in the ideal case (Fig. 12d), finite sections of such lines with upward magnetized dots exist, but the orientation of these fragments is chaotic. For an oblique square, such lines are almost absent, but "relicts" of the CAFM structure appear in the form of lines that contain magnetic dots with downward magnetic moments and are directed along the diagonals of the initial square lattice (Fig. 12b). Such lines are absent for the ideal structure and they are very slightly manifested for the circular array. The existence of these lines parallel to the borders of the system is evidently the manifestation of the geometry of the array borders in the magnetic structure of its central part. It is interesting that such a structure is closest to that observed experimentally (see Fig. 10 in [6]) for a field equal to almost half the saturation field. Unfortunately, data on the shape of the array are absent in that paper. For the circular system, where the effect of the symmetry of the system borders on its magnetic structure is expected to be minimal, the ideal structure shown in Fig. 12d is also not observed in the pure form. Only one important fragment of it, which is absent for square arrays, is present: the existence of pairs of magnetic dots with downward moments that are connected by the (2, 1) or (1, 2) translation vector, which represents the property of the ideal lattice at the level of the short-range order. We point to a higher symmetry of the state as a whole. Indeed, only pairs of such dots that are connected by one of these vectors, e.g., vector (2, 1) in Fig. 12d, are present for the ideal lattice, whereas pairs with both such vectors are present with the same probability for the finite array. We also emphasize that such elements are almost absent for the square geometry cases (Figs. 12a, 12b).

6. DISCUSSION AND CONCLUSIONS

In summary, it is reasonable to briefly discuss problems that remain beyond the scope of this paper. The numerical analysis of the problem of the transition from the out-of-plane (Ising) states to planar states when the magnetic field varies for various anisotropy constants β is of interest. In fact, this means the construction of a phase diagram on the (h, β) plane. Analysis of the zero field case shows [12] that, even for an extremely dense lattice of cylindrical magnetic dots made of soft magnetic materials, Ising states are stable for sufficiently large shape factors $L/R > 2$, where R and L are the radius and height of a dot. In fact, these states are stable when the anisotropy energy is comparable with the energy of interaction between neighboring dots. The experiment presented in [6] corroborates this rule.

It is rather difficult to theoretically analyze planar states. Our preliminary numerical data indicate that they correspond to complex noncollinear structures, which are characterized both by significant two-dimensional inhomogeneity with a scale of about the sample size, and by the presence of regions where neighboring magnetic moments are substantially noncollinear. The character of these structures is considerably determined by the shape of the array. It is evident that such states can be analyzed only numerically.

The effect of the violation of pure axial symmetry on Ising states analyzed above is also of interest. The simplest example of such a violation is the deviation of the external field from the given axis of the array (z axis). Preliminary numerical investigations of this case show that, for moderate values of the anisotropy constant $\beta = 5/a^3$ and for the field $h_z = 1$ along the symmetric direction, the inclusion of the transverse magnetic field $h = 0.2$ (field inclination by approximately 11°) gives rise to the appearance of small deviations (less than 1.5°) of magnetic moments from the given axis, but does not affect the out-of-plane structure of magnetic moments that is of interest. More detailed investigations of the role of anisotropy and the magnetic-field inclination and the properties of planar states, as well as a discussion concerning a very interesting question on the effect of the random spread of the parameters of single magnetic dots, are beyond the scope of this work.

Thus, the above results may be applied to the systems of magnetic dots with various properties of a single dot under the condition that magnetic moments are collinear to the given axis of the system. In this case, the anisotropy of a single dot is unimportant and the characteristic field $H_* = \mu_0/a^3$ is the only parameter important for comparison with experiment. The characteristic

instability fields studied here and in [14] range from $0.23H_*$ [see Eq. (12)] to $H_1 \approx 9H_*$. Let us discuss these fields for various systems, primarily those realized experimentally. We start with the case of dots in the homogeneous state. Consider magnetic dots in the form of a right circular cylinder with a diameter $D = 2R$ and height L made of a material with the saturation magnetization M_s. In this case, $\mu_0 = \pi R^2 L M_s$. For dense lattices, the diameter D of a dot is comparable with the lattice constant a and the thickness of the dot may significantly vary. For this reason, it is convenient to introduce the shape parameter $\lambda = L/R$ and write the characteristic field in the form

$$H_* = \pi\lambda M_s (R/a)^3. \quad (17)$$

The system of dots with $2R = L = 200$ nm made of dysprosium ($4\pi M_s = 34$ kOe) and organized in a not too dense square lattice with a period $a = 500$ nm was used in [7] and is characterized by a field $H_* = 200$ Oe. Magnetic dots made of soft magnetic materials such as permalloy constitute a more standard case. In these systems, perpendicular magnetization is realized for $\lambda \geq 2$ [12] and quite dense lattices with a up to $1.1 \times 2R$ are studied [1–3, 5, 6]. Detailed experimental data are presented in [6] for sufficiently large (several centimeters squared) lattices of magnetic dots with a period of 100–200 nm, which are made of various soft magnetic materials with perpendicular magnetization and have a diameter of 60–180 nm and $L/R \leq 6$. Even for lower saturation magnetization of soft magnetic materials (e.g., $4\pi M_s = 10$ and 6 kOe for permalloy and nickel, respectively), such systems may have higher H_* values reaching several kilo-oersteds due to the high lattice density and large λ value.

For magnetic dots in the vortex state made of standard materials such as permalloy for which M_s is high, the vortex core size does not exceed 15–20 nm. For this reason, H_* is quite low and is 30 Oe even for the optimum sizes $L = 50$ nm and $2R = a = 200$ nm. However, weak effective fields of remagnetization of single magnetic dots are a positive factor for their use in the logic elements of computers. The scheme of such a device was realized in [4] on the basis of a system of dots with a radius of less than 50 nm and planar magnetization.

Thus, our analysis reveals the considerable effect of borders on the character of the magnetization of finite arrays of dipole-coupled particles (magnetic dots). Differences in the behavior of a real system and idealized model of the infinite lattice are most substantial in the region of the destruction of the CAFM order. For a finite system, this destruction first occurs at the edges of the system, then near the extended regions of the surface, and only then expands to the entire interior of the array. In this case, the critical field at which this simplest state is destroyed is much weaker than the field characteristic for the ideal infinite system. In contrast to this scenario, when the field decreases, the destruction of the saturated state always begins far from the borders. By means of numerical analysis, we determine intermediate magnetic structures that are formed after the loss of the stability of the indicated simple phases. It turns out that magnetic structures in the intermediate field region, which are obtained when the field both increases and decreases, for finite arrays strongly differ from those characteristic of the ideal infinite system.

We showed that, in the presence of simple stacking faults such as single vacancies or their clusters in the lattice, localized deviations from the CAFM structure appear on these faults for quite weak fields. The magnetization curve of clusters of vacancies consists of several jumps with very characteristic relations between the fields and jump values. In essence, the section of such a step magnetization curve can be treated as an indicator of this magnetic defect. Therefore, analysis of the magnetization process for an array of magnetic dots for fields weaker than the instability field of the ideal CAFM structure can be used for diagnostics on the structure quality of series of samples. For comparison with experiment, it is important that all characteristic fields for both a finite array with the ideal lattice and lattice with vacancies are expressed in terms of the quantity H_* that is universal for a given system and is specified by Eq. (17). This characteristic field is determined by the lattice constant for dots and the material and magnetic structure of a single dot of the array, is primarily sensitive to the lattice density, and may vary from tens to thousands of oersteds for various systems.

REFERENCES

1. R. Skomski, J. Phys.: Condens. Matter **15**, R841 (2003).
2. M. Kläui, C. A. F. Vaz, L. Lopez-Diaz, and J. A. C. Bland, J. Phys.: Condens. Matter **15**, R985 (2003).
3. S. O. Demokritov, B. Hillebrands, and A. N. Slavin, Phys. Rep. **348**, 441 (2001).
4. R. P. Cowburn and M. E. Welland, Science **287**, 1466 (2000).
5. M. N. Baibich, J. M. Broto, A. Fert, *et al.*, Phys. Rev. Lett. **61**, 2472 (1988).
6. C. A. Ross, M. Hwang, M. Shima, *et al.*, Phys. Rev. B **65**, 144417 (2002).
7. P. D. Ye, D. Weiss, K. von Klitzing, *et al.*, Appl. Phys. Lett. **67**, 1441 (1995).
8. J. M. Luttinger and L. Tisza, Phys. Rev. **70**, 954 (1946).
9. P. I. Beloborodov, R. S. Gekht, and V. A. Ignatchenko, Zh. Éksp. Teor. Fiz. **84**, 1097 (1983) [Sov. Phys. JETP **57**, 636 (1983)].
10. J. G. Brankov and D. M. Danchev, Physica A (Amsterdam) **144**, 128 (1987); S. Prakash and C. L. Henley, Phys. Rev. B **42**, 6574 (1990).
11. K. Yu. Guslienko, Appl. Phys. Lett. **75**, 394 (1999).
12. K. Yu. Guslienko, S. Choe, and S. Shin, Appl. Phys. Lett. **76**, 3609 (2000).

13. S. V. Maleev, Zh. Éksp. Teor. Fiz. **70**, 2374 (1976) [Sov. Phys. JETP **43**, 1240 (1976)].
14. J. E. L. Bishop, A. Yu. Galkin, and B. A. Ivanov, Phys. Rev. B **65**, 174403 (2002).
15. M. Seul and R. Wolfe, Phys. Rev. Lett. **68**, 2460 (1992).
16. I. Booth, A. B. MacIsaac, J. P. Whitehead, and K. De'Bell, Phys. Rev. Lett. **75**, 950 (1995).
17. Ar. Abanov, V. Kalatsky, V. L. Pokrovsky, and W. M. Saslow, Phys. Rev. B **51**, 1023 (1995).
18. A. B. MacIsaac, J. P. Whitehead, M. C. Robinson, and K. De'Bell, Phys. Rev. B **51**, 16033 (1995).
19. J. Arlett, J. P. Whitehead, A. B. MacIsaac, and K. De'Bell, Phys. Rev. B **54**, 3394 (1996).
20. E. Y. Vedmedenko, A. Ghazali, and J.-C. S. Lévy, Surf. Sci. **402**–**404**, 391 (1998).
21. E. Y. Vedmedenko, H. P. Oepen, A. Ghazali, *et al.*, Phys. Rev. Lett. **84**, 5884 (2000).
22. E. Y. Vedmedenko, A. Ghazali, and J.-C. S. Lévy, Phys. Rev. B **59**, 3329 (1999).
23. W. Wernsdorfer, Adv. Chem. Phys. **118**, 99 (2001).
24. K. P. Belov, A. K. Zvezdin, A. M. Kadomtseva, and R. Z. Levitin, *Orientational Transitions in Rare-Earth Magnets* (Nauka, Moscow, 1980) [in Russian].
25. V. G. Bar'yakhtar and B. A. Ivanov, in *Soviet Science Reviews, Section A: Physics Reviews*, Ed. by I. M. Khalatnikov (Harwood Academic, Amsterdam, 1985), Vol. 6, p. 404.
26. D. L. Mills, Phys. Rev. Lett. **20**, 18 (1968).
27. F. Keffer and H. Chow, Phys. Rev. Lett. **31**, 1061 (1973).
28. W. E. Tennant, R. B. Bailey, and P. L. Richards, in *Proceedings of the Conference on Magnetism and Magnetic Materials, San Francisco, 1974*, Ed. by C. D. Graham, G. H. Lander, and J. J. Rhyne (American Inst. of Physics, New York, 1975), AIP Conf. Proc., No. 24.
29. R. W. Wang, D. L. Mills, E. E. Fullerton, *et al.*, Phys. Rev. Lett. **72**, 920 (1994).
30. R. W. Wang, D. L. Mills, E. E. Fullerton, *et al.*, Phys. Rev. B **53**, 2627 (1996).
31. S. Rakhmanova, D. L. Mills, and E. E. Fullerton, Phys. Rev. B **57**, 476 (1998).
32. N. A. Usov and S. E. Peschany, J. Magn. Magn. Mater. **118**, L290 (1993).
33. T. Shinjo, T. Okuno, R. Hassdorf, *et al.*, Science **289** (5481), 930 (2000); R. P. Cowburn, D. K. Koltsov, A. O. Adeyeye, *et al.*, Phys. Rev. Lett. **83**, 1042 (1999); K. Runge, T. Nozaki, U. Okami, *et al.*, J. Appl. Phys. **79**, 5075 (1996); A. Fernandez and C. J. Cerjan, J. Appl. Phys. **87**, 1395 (2000); T. Pokhil, D. Song, and J. Nowak, J. Appl. Phys. **87**, 6319 (2000).
34. A. I. Akhiezer, V. G. Bar'yakhtar, and S. V. Peletminskiĭ, *Spin Waves* (Nauka, Moscow, 1967; North-Holland, Amsterdam, 1968).
35. B. A. Ivanov, H. J. Schnitzer, F. G. Mertens, and G. M. Wysin, Phys. Rev. B **58**, 8464 (1998).
36. B. A. Ivanov, C. E. Zaspel, and A. Yu. Merkulov, J. Appl. Phys. **89**, 7198 (2001); Phys. Rev. B **68**, 212403 (2003).

Translated by R. Tyapaev

Journal of Experimental and Theoretical Physics, Vol. 101, No. 6, 2005, pp. 1122–1129.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 6, 2005, pp. 1279–1287.

ELECTRONIC PROPERTIES OF SOLIDS

Mechanisms of Low-Temperature High-Frequency Conductivity in Systems with a Dense Array of $Ge_{0.7}Si_{0.3}$ Quantum Dots in Silicon

I. L. Drichko[a], A. M. D'yakonov[a], I. Yu. Smirnov[a], A. V. Suslov[b], Yu. M. Gal'perin[a,c], A. I. Yakimov[d], and A. I. Nikiforov[d]

[a] *Ioffe Physicotechnical Institute, Russian Academy of Sciences, Politekhnicheskaya ul. 26, St. Petersburg, 194021 Russia*
e-mail: Irina.L.Drichko@mail.ioffe.ru
[b] *National High Magnetic Field Laboratory, Tallahassee, Florida 32310, USA*
[c] *Department of Physics and Center for Advanced Materials, University of Oslo, PO Box 1048 Blindern, 0316 Oslo, Norway*
[d] *Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia*

Received June 3, 2005

Abstract—High-frequency (HF) conductivity in systems with a dense (with a density of $n = 3 \times 10^{11}$ cm^{-2}) array of self-organized $Ge_{0.7}Si_{0.3}$ quantum dots in silicon with different boron concentrations n_B is determined by acoustic methods. The measurements of the absorption coefficient and the velocity of surface acoustic waves (SAWs) with frequencies of 30–300 MHz that interact with holes localized in quantum dots are carried out in magnetic fields of up to 18 T in the temperature interval from 1 to 20 K. Using one of the samples ($n_B = 8.2 \times 10^{11}$ cm^{-2}), it is shown that, at temperatures $T \le 4$ K, the HF conductivity is realized by the hopping of holes between the states localized in different quantum dots and can be explained within a two-site model in the case of $\omega\tau_0 \gg 1$, where ω is the SAW frequency and τ_0 is the relaxation time of the populations of the sites (quantum dots). For $T > 7$ K, the HF conductivity has an activation character associated with the diffusion over the states at the mobility threshold. In the interval 4 K $< T <$ 7 K, the HF conductivity is determined by a combination of the hopping and activation mechanisms. The contributions of these mechanisms are distinguished; it is found that the temperature dependence of the hopping HF conductivity approaches saturation at $T^* \approx 4.5$ K, which points to a transition to the regime of $\omega\tau_0 \le 1$. A value of $\tau_0(T^*) \approx 5 \times 10^{-9}$ s is determined from the condition $\omega\tau_0(T^*) \approx 1$. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Quantum dots represent a limiting case of low-dimensional systems in which the motion of particles is restricted in all three dimensions. Quantum dots are often called artificial atoms, because their electron spectra are discrete in spite of the fact that a real dot may consist of tens of thousands of atoms. The conductivity of dense arrays (ensembles) of quantum dots is different from zero even at very low temperatures. The transport properties of germanium quantum-dot ensembles in silicon δ-doped with boron with a density of $n = 3 \times 10^{11}$ cm^{-2} were studied in detail in [1]. The authors of these studies came to the conclusion that the conductivity at low temperatures is realized by the hopping of holes between the localized states of different quantum dots.

The complexity of the objects to be investigated and the ambiguity in the interpretation of their transport properties requires application of other experimental methods, for example, acoustic methods, that allow one to determine the high-frequency (HF) conductivity of a system in a contactless way. This method has already been used by the present authors, in particular, for investigating the HF conductivity in a system with pure germanium quantum dots in silicon at low temperatures [2]. These investigations have also led to the conclusion about the hopping mechanism of HF conductivity.

For the first time, the absorption of surface acoustic waves (SAWs) by a system with quantum dots was used in [3] for studying the properties of large (250–500-nm) quantum dots in *n*-GaAs/AlGaAs samples that were obtained by holographic lithography followed by ion etching. The authors of [3] interpreted their results as relaxation absorption associated with the transitions of the electrons localized in quantum dots between the energy levels within the same quantum dot [4].

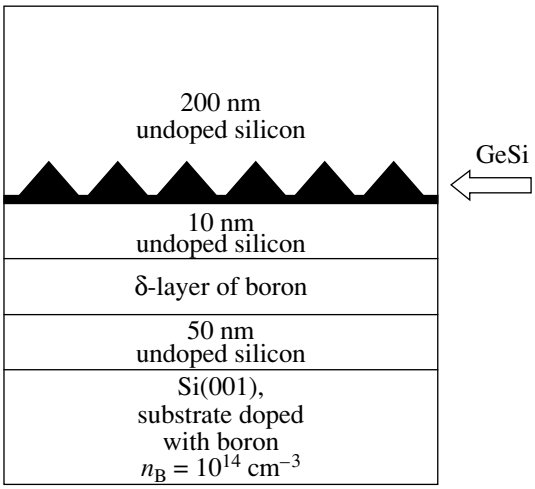


Fig. 1. Structure of samples.

In the present paper, we apply acoustic methods to investigate the HF conductivity of an ensemble of $Ge_{0.7}Si_{0.3}$ quantum dots in a silicon matrix. The composition was determined by the Raman scattering technique. The dc transport properties of such systems [5] differ from the properties of samples with pure germanium quantum dots [1]. This difference is likely to be associated with the fact that the potential well for holes in the region of $Ge_{0.7}Si_{0.3}$ dots is shallower than that in the case of pure germanium dots.

2. EXPERIMENTAL RESULTS

Acoustic methods consist in measuring the acoustoelectric characteristics—the absorption and the velocity of SAWs; this allows one to determine the HF conductivity of systems by a noncontact method.

Since germanium and silicon are not piezoelectric materials, we applied a hybrid technique, in which a SAW propagates along the surface of a piezodielectric $LiNbO_3$ plate, while a sample is slightly pressed by a spring to the surface of the $LiNbO_3$ plate. In this case, the electric field induced by the SAW and having the same frequency as the SAW penetrates into the sample; however, mechanically, the sample remains uncoupled [6]. The absorption of the SAW and the change in its velocity are determined by the HF conductivity of the sample under investigation.

The measurements were carried on four samples with a dense array ($n = 3 \times 10^{11}$ cm^{-2}) of self-organized $G_{0.7}Si_{0.3}$ quantum dots in a magnetic field of 18 T. The structure of samples is shown in Fig. 1. The layer of quantum dots lies at a depth of 2000 Å from the surface of a sample. The quantum dots have the shape of pyramids with a 120×120 Å^2 square base and a height of about 20 Å. The samples were δ doped with boron with a concentration of $n_B = 2.7 \times 10^{11}$ cm^{-2} and 6.8×10^{11} cm^{-2} (sample 1), 8.2×10^{11} cm^{-2} (sample 2), and 11×10^{11} cm^{-2} (sample 3). The measurements were carried out in the range of temperatures from 1 to 20 K with SAWs of frequencies from 30 to 300 MHz. A magnetic field was applied perpendicular to the plane in which the quantum dots were formed. The measurements of the sample with $n_B = 2.7 \times 10^{11}$ cm^{-2} did not reveal any acoustoelectric phenomena within the measurement accuracy.

Figure 2 represents the absorption $\Delta\Gamma \equiv \Gamma(H) - \Gamma(0)$ and the velocity $\Delta V/V \equiv [V(H) - V(0)]/V(0)$ of a SAW of frequency $f = 28$ MHz as a function of the magnetic field strength for sample 2 in the range of temperatures from 1 to 4 K. In high magnetic fields, these quantities approach saturation. It is worth noting that, as the magnetic field increases, the absorption decreases; i.e., $\Delta\Gamma < 0$. In magnetic fields of $H < 5$ T, both the absorption and the velocity of the SAW are proportional to H^2. For $H > 5$ T, the magnetic-field dependence of the

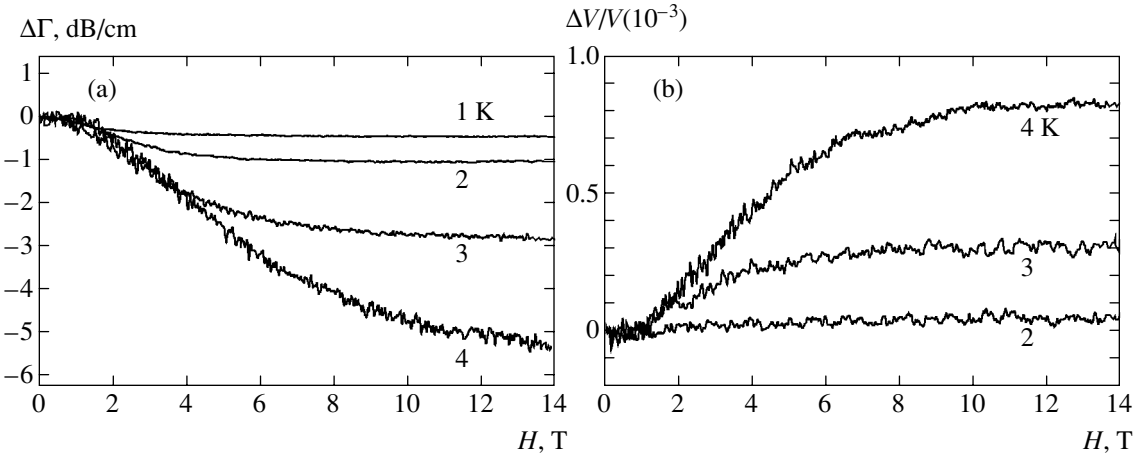


Fig. 2. (a) Absorption ΔΓ and (b) relative velocity variation $\Delta V/V$ of a SAW as a function of magnetic field H for sample 2 at T = 1–4 K and f = 28 MHz.

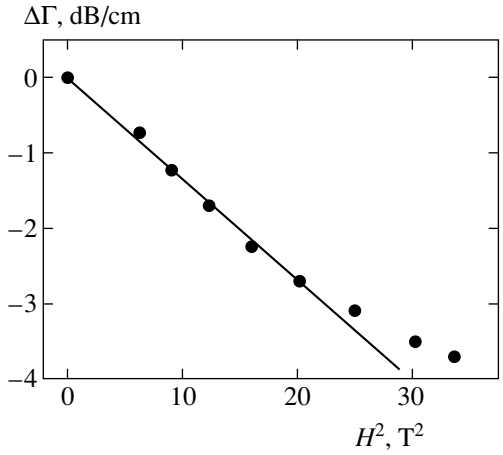


Fig. 3. The function $\Delta\Gamma(H^2)$ for sample 2 at $T = 4$ K and $f = 29.5$ MHz.

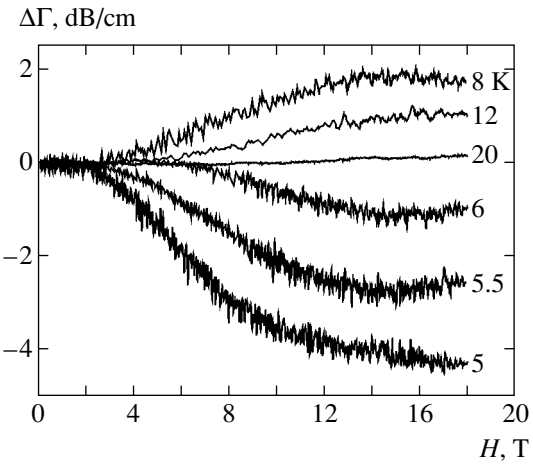


Fig. 4. The function $\Delta\Gamma(H)$ for sample 2 for $T = 5$–20 K and $f = 28$ MHz.

absorption $\Delta\Gamma$ deviates from the quadratic behavior as demonstrated in Fig. 3 by measurements carried out on sample 2 at $T = 4.2$ K and $f = 29.5$ MHz.

All the other samples exhibit similar behavior, but in different temperature intervals: the higher the boron doping level, the lower the temperatures at which the samples exhibit this behavior. Sample 1 exhibits such behavior up to 8 K, sample 2, up to 4.2 K, and sample 3, only up to 2.8 K.

Figure 4 represents the function $\Delta\Gamma(H)$ for sample 2 in the temperature interval 5–20 K and at a SAW frequency of 28 MHz. One can see that, when $T > 5$, the quantity $\Delta\Gamma(H)$ is negative and increases with temperature; when $T > 6$ K, it changes its sign.

As temperature increases, $\Delta\Gamma$ becomes positive in all the samples; in samples with higher concentration of boron, $\Delta\Gamma$ changes its sign at lower temperatures.

3. DISCUSSION OF EXPERIMENTAL RESULTS

3.1. Low Temperatures

At low temperatures (Fig. 2a), the absorption monotonically decreases as the magnetic field increases, $\Delta\Gamma(H) < 0$. Such behavior in $\Delta\Gamma(H)$ is typical of the hopping HF conductivity and is usually associated with a decrease in the overlap of the wavefunctions of electrons (holes) localized at the centers between which the hopping occurs. The function $\Delta\Gamma(H)$ was quantitatively analyzed in [7] within the model of hydrogen-like dot centers. The hopping character of the HF conductivity in the objects under investigation was also confirmed by dc measurements [5]; the results of these measurements were interpreted under the assumption that the hopping of holes occurs between the states that are localized in different quantum dots.

Within this model, the hopping absorption of SAWs by localized carriers in a strong magnetic field is described by the interpolation formula

$$\Delta\Gamma(H) = -\Gamma(0) + B/H^2$$

(B is a coefficient), which can rigorously be justified in the case of hydrogen-like centers [8]. This formula makes it clear that a strong transverse magnetic field suppresses the hopping absorption and that $\Delta\Gamma(H) = -\Gamma(0)$ as $H \longrightarrow \infty$. Thus, it becomes possible to determine the absorption $\Gamma(0)$ in the absence of a magnetic field. The function $\Delta\Gamma(H)$ in a strong magnetic field approaches saturation, which is clearly seen in sample 2 at very low temperatures (Fig. 2a). When the saturation could not be attained in the magnetic fields available in the experiment, we determined the value of $\Gamma(0)$, which corresponds to the intersection of the linear function $\Delta\Gamma(1/H^2)$ with the vertical axis as $1/H^2 \longrightarrow 0$; this situation is demonstrated in Fig. 5 for sample 1.

The simultaneous measurement of the absorption and the velocity of a SAW allows one to determine the complex HF conductivity

$$\sigma^{\mathrm{hf}} = \sigma_1(\omega) - i\sigma_2(\omega)$$

at a frequency of $\omega = 2\pi f$ [6]. As pointed out above, in the range of magnetic fields $H \longrightarrow \infty$, the parameters Γ and $\Delta V/V$ and the conductivities σ_1 and σ_2 in the absence of a magnetic field can be determined by the formulas

$$\frac{\Delta V}{V}(H \longrightarrow \infty) - \frac{\Delta V}{V}(0) = A\left[1 - \frac{1 + \Sigma_2(0)}{\Sigma_1^2(0) + [1 + \Sigma_2(0)]^2}\right],$$
$$\Gamma(0) = \frac{kA\Sigma_1(0)}{\Sigma_1^2(0) + [1 + \Sigma_2(0)]^2}. \quad (1)$$

Here,

$$A = 110.2b(k)\exp[-2k(a+d)],$$

$$\Sigma_i = \frac{4\pi\sigma_i}{\varepsilon_s V}t(k),$$

k and V are the wavevector and the velocity of a SAW, respectively; the functions $b(k)$ and $t(k)$ depend on k, the gap a between a sample and the lithium niobate plate, the depth d of the plane with quantum dots, and the dielectric constants of lithium niobate and vacuum; and ε_s is the dielectric constant of a sample [6].

One can solve this system of equations for $\Sigma_1(0)$ and $\Sigma_2(0)$ (and, hence, determine $\sigma_1(0, \omega)$ and $\sigma_2(0, \omega)$) only if the gap a between a sample and the surface of the lithium niobate plate is known. Unfortunately, the earlier developed method [6] cannot be directly applied to this case because it is based on the assumption that the conductivity does not depend on frequency. This assumption was justified for delocalized electrons [6] and is inapplicable to the study of dielectric samples with quantum dots. Therefore, we determined the gap by the number of Newton's rings produced at the contact of the sample and lithium niobate planes; for this purpose, we made a hole in the sample holder. This method is not very accurate; however, it makes it possible to evaluate the absolute values of the conductivities $\sigma_i(0, \omega)$ and to determine their dependence on the magnetic field, frequency, and temperature. In our experiment, the gap a varied from 0.3 to 0.7 μm, depending on the sample setup.

By solving the system of equations (1) with a known value of the gap a for $\Sigma_1(0)$ and $\Sigma_2(0)$, we determined $\sigma_1(0, \omega)$ and $\sigma_2(0, \omega)$ for various SAW frequencies and temperatures. The conductivity components σ_1 and σ_2 proved to be independent of frequency to within 15 and 25%, respectively. The temperature dependence of $\sigma_1(0)$ and $\sigma_2(0)$ for sample 2 at low temperatures is shown in Fig. 6; one can see that $\sigma_1 > \sigma_2$ at these temperatures. In all the samples, the temperature dependence of the real part of conductivity is well described by the power law $\sigma_1^{(i)} = B_i T^{2.4}$, where B_i is the proportionality factor and i is a sample number. To illustrate this fact, in the inset to Fig. 6, we plot $\sigma_1^{(i)}/B_i$ as a function of $T^{2.4}$ for different samples. One can see that all points lie on the same straight line to a good degree of accuracy. The power-law dependence of the HF conductivity $\sigma_1(0)$ on temperature and its independence of frequency qualitatively agree with the predictions of the two-site model of hopping HF conductivity under the assumption that the period of a SAW is much smaller than the typical relaxation time τ_0 of the population of sites [7, 8]. According to this model, the absorption of a SAW and the variation of its velocity are

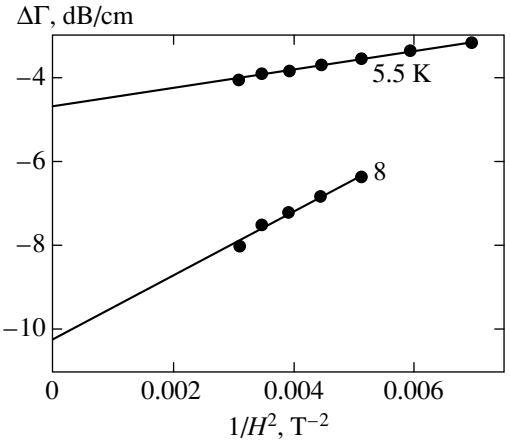


Fig. 5. The function $\Delta\Gamma(1/H^2)$ for sample 1 for temperatures of $T_1 = 5.5$ K and $T_2 = 8$ K and a SAW frequency of $f = 143$ MHz.

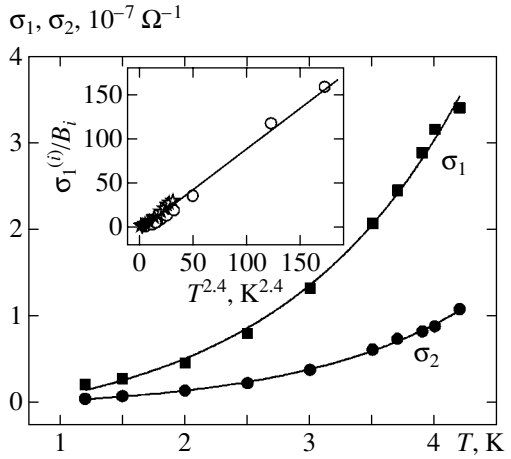


Fig. 6. $\sigma_1(0)$ and $\sigma_2(0)$ versus temperature for sample 2. The inset represents $\sigma_1^{(i)}/B_i$ as a function of $T^{2.4}$ for samples 1 (○), 2 (★), and 3 (▼).

associated with the absorption and emission of phonons during the relaxation of the populations of two-site clusters to their adiabatically equilibrium values. When $\omega\tau_0 \gg 1$, the real part of conductivity must be proportional to $1/\tau_0 \propto T^\alpha$, where the exponent α depends on the interaction mechanism. Thus, if the hopping conductivity is dominant, then the temperature behavior of σ_1 must be the same for different samples, as is demonstrated in the inset to Fig. 6.

Another prediction of the two-site model is a transition to a temperature-independent absorption at low frequencies, where the condition $\omega\tau_0 \ll 1$ is fulfilled. In this regime, absorption must be proportional to frequency. In spite of relatively low frequencies of the SAWs, such a regime has not been directly observed at low temperatures.

If we assume that all the holes from boron atoms populate the states in the quantum dots, then there are, on average, 2.3 holes per dot in sample 1, 2.7 in sample 2, and 3.7 in sample 3. This means that, in samples 1 and 2, two lower lying states are predominantly occupied, while the third excited state is occupied only in some of the dots.

It is natural to assume that the hops (in which phonons take part) between dots with occupied and empty states may lead to the relaxation absorption like in macroscopically homogeneous doped semiconductors. In this case, the dc conductivity corresponds to a percolation cluster constructed from localized states of quantum dots. Therefore, the dc conductivity must decay exponentially with temperature, as observed in sample 2 [5].

According to the two-site model, which is based on hydrogen-like localized states, the transition from the dependence $\Delta\Gamma \propto H^2$ in a weak magnetic field to $\Delta\Gamma \propto H^{-2}$ in a strong field occurs at $a_H \approx \xi$, where $a_H = \sqrt{c\hbar/eH}$ is the magnetic length and ξ is the decay length of a localized state [9]. This formula allows one to evaluate the decay length ξ. Unfortunately, this procedure cannot be quantitatively applied to the present situation because the effective attenuation length ξ_{eff} in a system consisting of granules may be substantially renormalized (see, for example, [10–13]). In a typical situation, the effective attenuation length increases compared with the attenuation length in the intergranular region, and reliable estimates can be made only for $\xi_{\text{eff}} \gg l \gg \xi$, where l is the intergranular distance. If we apply the two-site model with renormalized parameters to study a system of quantum dots, then we obtain $\xi_{\text{eff}} =$ 80 Å for sample 1 and $\xi_{\text{eff}} = 120$ Å for samples 2 and 3. Since the strong inequalities presented above are not satisfied, these estimates must be considered as merely tentative ones. Apparently, the strong inequality $\sigma_2 \gg \sigma_1$, which follows from the theory of hopping HF conductivity for hydrogen-like dot centers, is not satisfied for the same reason.

3.2. High Temperatures

The figures presented above show that, in strong magnetic fields, the quantity $|\Delta\Gamma(H)|$ increases up to $T = 4.2$ K (see Fig. 2a); for $T > 4.2$ K, it starts to decrease (Fig. 4). If the hopping mechanism of conductivity, which was assumed to be dominant at low temperatures (1–4.2 K), was the only mechanism, then, according to theoretical predictions, the condition $\omega\tau_0 \ll 1$ would hold at sufficiently high temperatures and the temperature dependence would approach saturation. Since there are no such phenomena in the experiment, it is natural to assume that there is another conductivity mechanism that plays the dominant role at high temperatures.

We assume that, at high temperatures, the main role is played by the activation of carriers to the mobility threshold. Below, we will show that this assumption is self-consistent. Since the imaginary part of the HF conductivity in the region of diffusion at the percolation level is small, formulas (1) for an arbitrary magnetic field can be substantially simplified:

$$\gamma(H) = \frac{\Gamma(H) - \Gamma(0)}{kA} = \frac{\Sigma_1(H)}{1 + \Sigma_1^2(H)} - \frac{\Sigma_1(0)}{1 + \Sigma_1^2(0)},$$

$$\nu(H) = \frac{V(H) - V(0)}{V(0)A} = \frac{1}{1 + \Sigma_1^2(H)} - \frac{1}{1 + \Sigma_1^2(0)}.$$

This yields the following useful relation:

$$\Sigma_1(0) = \frac{\eta(H)\Sigma_1(H) + 1}{\Sigma_1(H) - \eta(H)},$$

where $\eta(H) = \gamma(H)/\nu(H)$. In particular, if the experimentally measured absorption $\Gamma(H)$ has a maximum at a certain value of magnetic field $H = H_{\max}$, then $\Sigma_1(H_{\max}) = 1$, and $\Sigma_1(0)$ is determined from the measured value of $\eta(H_{\max})$ as

$$\Sigma_1(0) = \frac{1 + \eta(H_{\max})}{1 - \eta(H_{\max})}. \qquad (2)$$

It is this method by which the HF conductivity was determined in the interval of temperatures $T = 7$–8 K, where the function $\Gamma(H)$ attains its maximum. The conductivity $\Sigma_1(0)$ at $T = 12$ K was determined by a formula of type (1) under the condition $\Sigma_1^2(0) \gg 1$. It turns out that the temperature dependence of conductivity has the form $\sigma_1(0) \propto \exp(-E_a/k_B T)$, where E_a is the activation energy; i.e., $\sigma_1(0)$ obeys the activation law. However, this exponential dependence was measured in a very narrow temperature interval. To extend this interval, we chose the value of σ_1 at $T = 4.2$ K as the fourth point, so that this value does not exceed 20% of the hopping conductivity at the given temperature. From this dependence, the activation energy $E_a(H = 0)$ was found to be 2.5 meV.

The exponential behavior of the conductivity as a function of $1/T$ and its large absolute value, $\sigma_1(0) \sim 10^{-5}\ \Omega^{-1}$ for $T > 12$ K, confirm our hypothesis that the absorption is attributed to the diffusion at the mobility threshold. The results thus obtained, together with the low temperature data, are represented by dots in Fig. 7.

In the range of intermediate temperatures (4–7 K), the HF conductivity is determined by a sum of the contribution of hopping between localized states in different quantum dots and the contribution of diffusion at the percolation level. The latter contribution depends on temperature by an activation law and should slightly

differ from the experimentally observed static conductivity. Interpolating the experimental results from the low- and high-temperature regions by a smooth curve (Fig. 7) and subtracting from this curve the result obtained by extrapolating the high temperature data, we obtain the temperature dependence of the purely hopping contribution to the conductivity. This dependence is shown in Fig. 8. One can see that the temperature dependence of the hopping contribution approaches saturation, which is in qualitative agreement with the predictions of the renormalized two-site model. If the saturation region starts at a temperature of $T^* \approx 4.5$ K, as in Fig. 8, then the condition $\omega\tau_0(T^*) \approx 1$ is fulfilled, where $\omega = 2\pi f$ ($f = 30$ MHz); hence, $\tau_0(T^*) \approx 5 \times 10^{-9}$ s.

3.3. Nonlinear Regime

We investigated the dependence of $\Delta\Gamma$ and $\Delta V/V$ on a magnetic field for various intensities of SAWs.

Figure 9 represents $\Delta\Gamma = \Gamma(H) - \Gamma(0)$ as a function of the magnetic field strength for various intensities of SAWs for sample 2. The SAW frequency was 28 MHz; the measurements were carried out at $T = 4.2$ K. When the SAW intensity at the input of a sample increased (the output power of the HF generator increased by 23 dB), $\Delta\Gamma$ was independent of the SAW intensity W; with a further increase of W, $|\Delta\Gamma|$ started to decrease in absolute value and changed its sign at a certain value of W. The quantity $\Delta V/V$ was virtually independent of W.

Within the model that accounts for the temperature dependence of $\Delta\Gamma(H)$, one may assume that the contribution of delocalized charge carriers to the conductivity increases with the SAW intensity.

Figure 10 represents $\Delta\Gamma$ as a function of the signal attenuation at the output of the HF generator for a magnetic field strength of $H = 5.8$ T. It is worth noting that this effect is of threshold character. According to the preliminary analysis, the threshold value of the electric field $\mathscr{E}_c$ induced by the SAWs is roughly proportional to the SAW frequency and weakly depends on the magnetic field. The presence of the threshold and the proportionality of the threshold field to the SAW frequency qualitatively agree with the assumption that the nonlinearity mechanism is attributed to the so-called impurity breakdown—the tunneling of holes to the mobility threshold in the inhomogeneous field of the acoustic wave. This nonlinearity mechanism was observed in bulk InSb samples [14]. Let us estimate the threshold electric field $\mathscr{E}_c$ for this mechanism:

$$\mathscr{E}_c = kE_a/2e,$$

where e is the electron charge. For $E_a \approx 2.5$ meV and $f = 30$ MHz, such an estimate yields a value of $\mathscr{E}_c \approx 1$ V/cm.

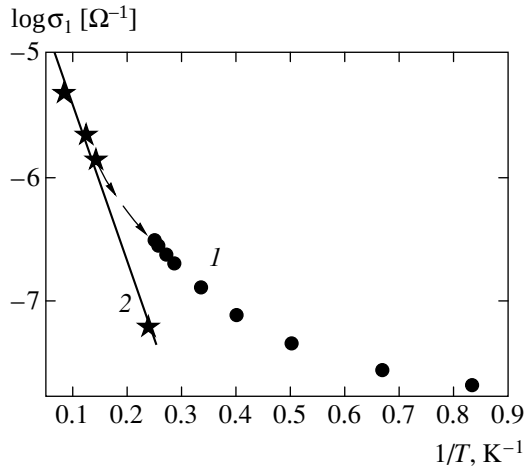


Fig. 7. $\sigma_1(0)$ as a function of $1/T$ in the temperature intervals (*1*) $T = 1$–4 K and (*2*) $T = 7$–12 K.

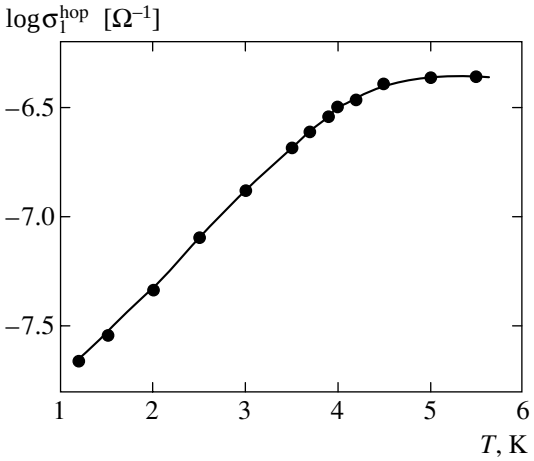


Fig. 8. The hopping contribution σ_1^{hop} to conductivity as a function of temperature.

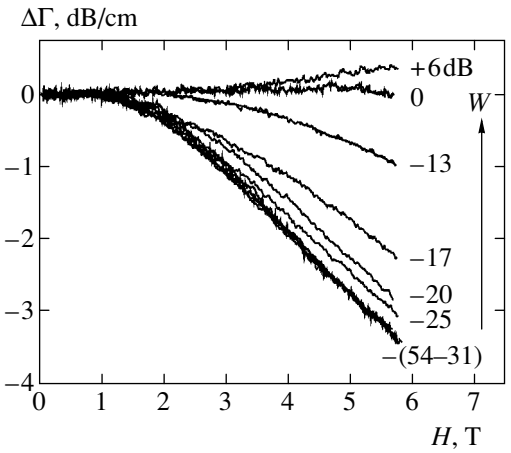


Fig. 9. The functions $\Delta\Gamma = \Gamma(H) - \Gamma(0)$ versus a magnetic field for various intensities of a SAW for sample 2 at $T = 4.2$ K and $f = 28$ MHz. The arrow points to the increase of the SAW intensity. The numbers at the curves indicate the values of attenuation at the output of the RF source that generates SAWs.

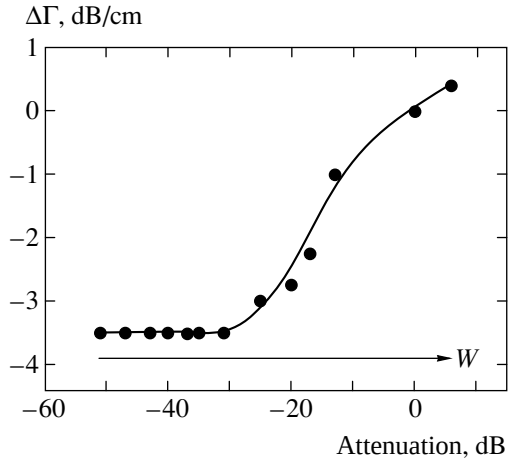


Fig. 10. $\Delta\Gamma$ as a function of the attenuation value at the output of the HF generator for sample 2 at $H = 5.8$ T and $T = 4.2$ K. The arrow points to the increase of the SAW power.

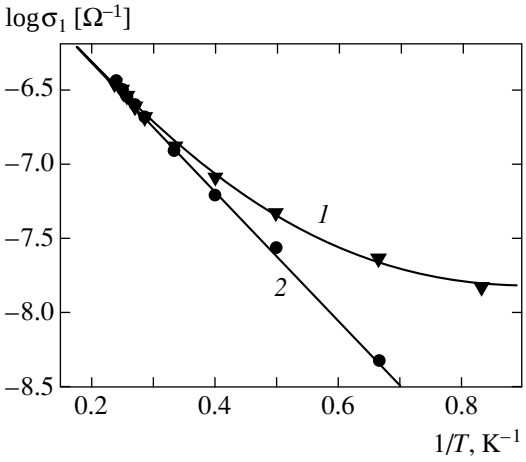


Fig. 11. σ_1^{hf} (*1*) and σ_1^{dc} (*2*) versus $1/T$ for sample 2 at $H = 0$ and $f = 28$ MHz.

The electric field induced by a SAW can be determined by the following formula [15]:

$$|\mathscr{E}|^2 = K^2 \frac{32\pi}{V}(\varepsilon_1 + \varepsilon_0) \times \frac{k\exp[-2k(a+d)]}{b_1^2(k)[\Sigma_1^2(0) + [1+\Sigma_2(0)]^2]} W, \quad (3)$$

where K^2 is the electromechanical coupling constant for $LiNbO_3$; W is the SAW power at the input of a sample divided by the width of the acoustic beam (which is equal to the aperture of the interdigital transducers that generate a SAW in lithium niobate); and $b_1(k)$ is a function that depends on the gap a, depth d of the layer with quantum dots, and the dielectric constants of lithium (ε_1), vacuum (ε_0), and a sample (ε_s).

However, the electric field induced by a SAW at the input of a sample is very roughly determined in the experiment because, in addition to the transducer loss, one should take into account the loss in the transmission line, which is not matched to a sample. Therefore, the power at the input of a sample is measured to within an order of magnitude. Our experimental estimate of the threshold field by formula (3) yields $\mathscr{E}_c \approx 4$ V/cm, which, with regard to the aforesaid, does not contradict the assumption about the impurity breakdown. However, the unambiguous identification of the nonlinearity mechanism requires additional investigations.

4. CONCLUSIONS

The complex HF conductivity of samples with a self-organized lattice of $Ge_{0.7}Si_{0.3}$ quantum dots in silicon has been determined by acoustic methods. We have shown that, in the low-temperature region ($T \le 4$ K), the HF conductivity is determined by the hopping over several quantum dots and is reasonably interpreted in terms of the two-site model with a renormalized localization length [9–12]. At extremely low temperatures, the characteristic relaxation time τ_0 is greater than the period of the acoustic wave. This leads to a weak (logarithmic) frequency dependence and power-law temperature dependence of the HF conductivity. As temperature increases, the relaxation time decreases, and, at $T \approx 4.5$ K, the temperature dependence of the hopping contribution approaches saturation, which is attributed to the transition to the regime of $\omega\tau_0 \le 1$, where the two-site model predicts very weak (logarithmic) temperature dependence and allows one to evaluate the time $\tau_0(4.5\text{ K}) \approx 5 \times 10^{-9}$ s. As temperature increases, there arises an additional contribution to the HF conductivity that depends on temperature by the activation law. As temperature increases further, this mechanism becomes the principal one and is interpreted as a diffusion over the states at the mobility threshold. Figure 11 represents, for comparison, the dc conductivity (σ_1^{dc}) measured in a sample whose properties are close to those of sample 2, and the HF conductivity (σ_1^{hf} determined from the acoustic measurements on sample 2, as functions of temperature. One can see that these curves diverge at low temperatures; moreover, $\sigma_1^{dc} < \sigma_1^{hf}$, which is characteristic of the region of hopping conductivity. The curves approach each other in the range of temperatures where the conductivity through delocalized states seems to appear on the scene; in this case, the HF conductivity should not differ from the static conductivity.

In [5], the temperature dependence of magnetoresistance in samples 1 and 2 was studied in magnetic fields of up to 6 T. The activation behavior of the resistance at low temperatures in different magnetic fields was explained within the model of multiparticle correla-

tions of holes localized in quantum dots, which lead to the formation of electron polarons in a disordered two-dimensional system. However, the results of the present paper do not give any grounds for this assumption, because they are explained by simpler arguments. The observed temperature dependence of the electric conductivity for $T > 5$ K was interpreted in [5] as a manifestation of variable-range hopping conductivity. In our view, the absolute value of the conductivity (greater than 10^{-6} Ω^{-1}) is too large for the realization of this mechanism.

Thus, in our view, the hopping mechanism of conductivity in Si/Ge systems with a dense array of self-organized $Ge_{0.7}Si_{0.3}$ quantum dots ($n = 3 \times 10^{11}$ cm^{-2}) is changed, as temperature increases, to the conductivity through delocalized states. We could distinguish these mechanisms and investigate their characteristics as functions of temperature and a magnetic field.

Note that the nature of states at the percolation level in the material investigated is not quite clear. Apparently, the high density of quantum dots and their large size may lead to a significant overlap of the wavefunctions of individual quantum dots, as well as to the strong Coulomb interaction of carriers in different dots. It is likely that the delocalized states are formed by the impurity band of boron in silicon; however, the nature of delocalized states in these systems requires additional analysis.

ACKNOWLEDGMENTS

We are grateful to G.M. Min'kov for measuring the dc transport properties of the samples and useful discussions, to V.A. Volodin for analyzing the composition of germanium nanoclusters by the Raman scattering technique, and to A.K. Gutakovskiĭ for investigating the structure by high-resolution electron microscopy techniques.

This work was supported by the Russian Foundation for Basic Research (project nos. 04-02-16246 and 03-02-16526). Part of the research was carried out at the National High Magnetic Field Laboratory, Tallahassee, Florida, and was supported by the National Science Foundation (cooperative agreement no. DMR-0084173) and by the State of Florida. The ultrasonic measurements in this laboratory were carried out within an In-House Research Program.

REFERENCES

1. A. I. Yakimov, A. V. Dvurechenskii, V. V. Kirienko, *et al.*, J. Phys.: Condens. Matter **11**, 9715 (1999); Phys. Rev. B **61**, 10868 (2000).
2. I. L. Drichko, A. M. Diakonov, V. I. Kozub, *et al.*, in *Proceedings of 11th International Symposium on Nanostructure: Physics and Technology* (Ioffe Physicotechnical Inst., St. Petersburg, 2003), p. 133.
3. G. R. Nash, S. J. Bending, M. Boero, *et al.*, Phys. Rev. B **59**, 7649 (1999).
4. A. Knäbchen, O. Entin-Vohlman, Y. M. Galperin, and Y. B. Levinson, Europhys. Lett. **39**, 419 (1997).
5. A. I. Yakimov, A. V. Dvurechenskii, G. M. Min'kov, *et al.*, Zh. Éksp. Teor. Fiz. **127**, 817 (2005) [JETP **100**, 722 (2005)].
6. I. L. Drichko, A. M. Diakonov, I. Yu. Smirnov, *et al.*, Phys. Rev. B **62**, 7470 (2000).
7. A. L. Efros and B. I. Shklovskii, in *Electron–Electron Interactions in Disordered Systems*, Ed. by A. L. Efros and M. Pollak (North-Holland, Amsterdam, 1985), p. 409; Y. M. Galperin, V. L. Gurevich, and D. A. Parshin, in *Hopping Transport in Solids*, Ed. by B. Shklovskii and M. Pollak (Elsevier, New York, 1991), p. 453.
8. Yu. M. Gal'perin and É. Ya. Priev, Fiz. Tverd. Tela (Leningrad) **28**, 692 (1986) [Sov. Phys. Solid State **28**, 385 (1986)].
9. B. I. Shklovskii and A. L. Éfros, *Electronic Properties of Doped Semiconductors* (Nauka, Moscow, 1979; Springer, New York, 1984), p. 416.
10. J. Zhang and B. I. Shklovskii, Phys. Rev. B **70**, 115317 (2004).
11. I. S. Beloborodov, A. V. Lopatin, V. M. Vinokur, and V. I. Kozub, cond-mat/0501094.
12. M. V. Feigel'man and A. S. Ioselevich, Pis'ma Zh. Éksp. Teor. Fiz. **81**, 341 (2005) [JETP Lett. **81**, 277 (2005)].
13. V. I. Kozub, V. M. Kozhevin, D. A. Yavsin, and S. A. Gurevich, Pis'ma Zh. Éksp. Teor. Fiz. **81**, 287 (2005) [JETP Lett. **81**, 226 (2005)].
14. Yu. M. Gal'perin, I. L. Drichko, and L. B. Litvak-Gorskaya, Fiz. Tverd. Tela (Leningrad) **28**, 3374 (1986) [Sov. Phys. Solid State **28**, 1899 (1986)].
15. I. L. Drichko, A. M. D'yakonov, I. Yu. Smirnov, and A. I. Toropov, Fiz. Tekh. Poluprovodn. (St. Petersburg) **34**, 436 (2000) [Semiconductors **34**, 422 (2000)].

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 101, No. 6, 2005, pp. 1130–1139.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 6, 2005, pp. 1288–1298.

ELECTRONIC PROPERTIES OF SOLIDS

Deformation of a Charge Density Wave in the Vicinity of a Point Contact with a Normal Metal

A. A. Sinchenko[a] and V. Ya. Pokrovskiĭ[b]

[a]*Moscow Engineering Physics Institute, Moscow, 115409 Russia*
[b]*Institute of Radio Engineering and Electronics, Russian Academy of Sciences, Moscow, 103907 Russia*
e-mail: pok@cplire.ru

Received June 17, 2005

Abstract—The current–voltage characteristics of Cu–$K_{0.3}MoO_3$ point contacts between a metal and a semiconductor with a charge density wave (CDW) are studied for various diameters of the contacts in a wide range of temperatures T and voltages V. In the interval 80 K $\lesssim T \lesssim$ 150 K, the current–voltage characteristics are correctly described in the framework of a semiconductor model: screening of an external electric field causes CDW deformation, shifts the chemical potential of quasiparticles, and changes the point contact resistance. It is shown that the chemical potential is above the middle of the Peierls gap in equilibrium and approaches the middle upon an increase in temperature. The current–voltage characteristics of point contacts with a diameter $d \gtrsim 100$ Å exhibit a sharp decrease in resistance for $|V| > V_t$, which is associated with the beginning of local CDW sliding within the contact region. The $V_t(d, T)$ dependence can be explained by the size effect in the CDW phase slip.

1. INTRODUCTION

A charge density wave (CDW) emerges in a low-dimensional conductor upon a decrease in temperature as a result of the Peierls transition. In this case, an energy gap 2Δ is formed on the Fermi surface, as a result of which the conductor passes to a semimetallic ($NbSe_3$) or semiconducting (blue bronze $K_{0.3}MoO_3$ and TaS_3) state [1]. The emerging state is characterized by a periodic modulation of electron density $\rho = \rho_0 + \rho_1\cos(qx + \phi)$, where ρ_0 and ρ_1 are the mean value and amplitude of electron density modulation, $q = 2k_F$ is the wavevector of the CDW, and ϕ is the CDW phase. Three-dimensional CDW ordering allows us to treat it as an electronic crystal. In the Peierls state at a finite temperature, single-electron excitations (electrons and holes) always exist and determine the conductivity and other kinetic properties of the system in electric fields below the critical value (i.e., when a CDW cannot move in the crystal as a single entity). Allowance for the effect of single-electron excitations is most important for CDWs with complete dielectrization of the electron spectrum. In this case, a Peierls conductor possesses physical properties analogous to those of a semiconductor. However, considerable differences between the properties of semiconductors with CDWs and conventional semiconductors exist even in weak electric fields. For example, any change in the CDW wavevector is associated with a change in the CDW charge and, hence (in view of electroneutrality), with a change in the balance between thermal single-electron excitations [2]. This, in turn, may lead to a considerable change in the linear conductivity ensured by these excitations at a low temperature. Thus, a change in the CDW wavevector plays the same role in semiconductors with CDWs as doping in conventional semiconductors—it changes the concentration of electrons and holes. The concept of a semiconductor with CDWs as an ordinary semiconductor with a doping level depending on external perturbation (temperature T or electric field E) was used in [3, 4] for describing the thermopower and conductivity of TaS_3.

The development of these ideas led to the construction of a semiconductor model of CDWs [5]. The essence of this model is that transport properties for a stationary (pinned) CDW are determined by electron and hole excitations over the Peierls gap 2Δ, while the difference $p - n$ in the linear concentrations of quasiparticles changes upon a change in the CDW wavevector q: $q - q_0 = (p - n)/\pi$, where q_0 is the value of q at $T = 0$. It was shown in the framework of this model that the temperature dependence of conductivity in unipolar semiconductors with CDWs is mainly determined by the $q(T)$ dependence, $q(T) - q_0 \propto \exp(-\Delta/kT)$, where k is the Boltzmann constant. The behavior of the thermopower, Hall effect, and temperature hysteresis in conductivity was adequately explained for various compounds in terms of the shift in the chemical potential ζ relative to the middle of the band gap. The chemical potential shift $\delta\zeta$ is in one-to-one correspondence with the CDW deformation δq, i.e., with deviation of q from the equilibrium value caused by a change in temperature or electric field.

The application of an electric field smaller than or on the order of the threshold field E_t to a Peierls semi-

conductor in the vicinity of contacts with a conducting metallic leads results in the extension or compression of the CDW. Upon a further increase in the field, the deformation attains its critical value, the CDW is ruptured and generated or annihilated over a CDW period; i.e., phase slip centers are formed and the CDW begins its motion.

In traditional geometry for measuring transport properties, the area of contact leads made of a normal metal is comparable to the transverse size of samples with CDWs; in this case, the critical shift in the chemical potential is as a rule on the order of a few millielectronvolts (i.e., is much smaller than the value of kT practically at all temperatures). The conversion of the current of normal charge carriers into the CDW current in this geometry was studied in detail in [6, 7]. Measurements of the spatial variation of the CDW wavevector showed that the CDW deformation in the contact region exists on the macroscopic scale of lengths. It was shown in our previous publication [8] that the situation changes qualitatively if the metal–CDW contact is of microscopic size. In this case, the electric field is mainly localized in the contact region with a size on the order of the contact diameter [9]. A very large CDW deformation and, accordingly, a considerable shift in the chemical potential for a stationary CDW can be attained. The proposed technique makes it possible in principle to determine the local position of the chemical potential level at various temperatures.

Here, we report on the results of a detailed experimental study of the characteristics of Cu–$K_{0.3}MoO_3$ point contacts in a temperature range of 77–200 K. The experimental results are used for determining the temperature dependences of the position of the chemical potential and the electric field screening coefficient near the contact between the normal metal and blue bronze. We also show that a transition to the local sliding of a CDW can be observed upon an increase in the contact diameter. The behavior of the threshold characteristics of CDWs in this case can be considered from the standpoint of the existence of a peculiar size effect associated with the fact that the phase slip conditions are determined by the diameter of the point contact.

In Section 2, physical processes in the vicinity of a point contact are described in the framework of the semiconductor model. The experimental technique is described in Section 3. The results of measurements of the temperature evolution in high-resistance Cu–$K_{0.3}MoO_3$ contacts are described in Section 4. In Section 5, the results of analogous studies for low-resistance contacts are considered. The effect of the contact diameter on the characteristics of a Peierls conductor in the vicinity of the contact with a normal metal are also discussed. In Section 6, the results are considered in the framework of the semiconductor model of CDWs.

2. THEORETICAL MODEL

It was mentioned above that the electric field is localized in a small region in the vicinity of a point contact [9]. If both materials are isotropic, the size of this region is approximately equal to the contact diameter. In our case of the contact between a normal metal and an anisotropic semiconductor with CDWs, the depth of this region is larger by a factor of $\sqrt{A}$, where A is the conductivity anisotropy (of the order of 10^3 for $K_{0.3}MoO_3$ [1]) in the case when the current flows along conducting chains. Let us consider the processes occurring in such a region for a contact between a metal (Cu in our case) and blue bronze ($K_{0.3}MoO_3$) upon the application of a positive voltage. The results obtained in [8] show that the CDW deformation δq in the vicinity of the point contact under the action of the electric field disturbs the balance between electrons and holes (i.e., leads to a change in the concentration difference $p - n$). Consequently, a chemical potential shift $\delta\zeta \approx \delta q(d\zeta/dq)$ emerges in this case. As a result, the contact resistance changes.

It is known from the measurements of thermopower and the Hall effect [10] that blue bronze exhibits n-type conductivity in the Peierls state. This means that the chemical potential in equilibrium is shifted above the effective middle of the Peierls energy gap determined from the condition $\mu_n n = \mu_p p$, where μ_n and μ_p are the mobilities of electron and hole excitations, respectively.

Figure 1 shows the energy band diagram in the vicinity of a point contact. The positive sign of the bias voltage V corresponds to the downward shift of the chemical potential (we assume that $\delta\zeta > 0$). In this case, the maximum of the contact resistance corresponds to the position of the chemical potential at the effective

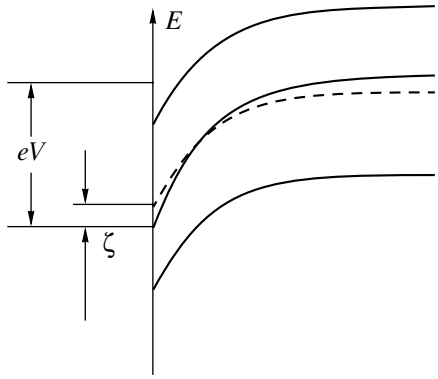


Fig. 1. Diagram illustrating distortion of energy bands in the vicinity of a metal–CDW (n-type semiconductor) point contact for a positive bias voltage. The dashed curve marks the effective middle of the Peierls gap (electrostatic potential); ζ is the chemical potential measured from the middle of the gap. In the vicinity of the contact, the electric field is partly screened: the slope of the electrostatic potential is smaller that of the electrochemical potential (middle solid curve). The electrochemical potential intersects the middle of the gap (transition from the n-type to p-type conductivity).

middle of the gap. Upon a further increase in the voltage, a transition from electron to hole conductivity takes place. Indeed, it can be seen from Fig. 1 that the CDW deformation near the contact occurs in such a way that the electric field determined by the electrostatic potential gradient (dashed curve in Fig. 1) decreases in the vicinity of the contact: the CDW is deformed under the action of this electric field until the field vanishes or its action is compensated by pinning forces.

For an appreciable CDW deformation, the chemical potential may be located below the middle of the gap, which corresponds to the hole-type conductivity. The dependence of the contact resistance R on the bias voltage is asymmetric with a peak shifted to the region of positive voltages. The transition through the peak is precisely the transition to the hole-type conductivity. To obtain a quantitative description of the $R(V)$ dependence, we must take into account the fact that the change in the resistance of the contact region occurs nonuniformly (see Fig. 1). For simplicity, we present R as the sum $R(\delta\zeta) + R_0$, where R_0 is independent of V, while $R(\delta\zeta)$ is determined by the conventional formula for the conductivity of quasiparticles ($\mu_n n + \mu_p p$). We assume that quantity $\delta\zeta$ is proportional to the voltage ($\delta\zeta = bV$). The case when $b = 1$ corresponds to complete screening of the applied field, i.e., the absence of bending in the energy bands depicted in Fig. 1. In other words, we represent the region of the voltage drop as the series connection of two subcircuits (with a uniform shift of the chemical potential and with a resistance independent of V):

$$R(V) = \frac{R_1}{\cosh[b(V - V_0)/kT]} + R_0. \qquad (1)$$

This is an expression with four fitting parameters: V_0 corresponds to the maximal resistance (the chemical potential is at the effective middle of the gap), while quantity bV_0 gives an estimate of the initial shift ζ_0 in the chemical potential relative to the effective middle of the gap. The value of R_1 is several times higher than R_0 as a rule; i.e., the CDW is deformed practically in the entire region of the contact voltage drop. Formula (1) was successfully used in [8] for describing the $R(V)$ dependences for Cu–$K_{0.3}MoO_3$ point contacts at T = 77 K. The properties of Cu–TaS_3 point contacts were studied and analyzed analogously [11]; voltage V_0 turned out to be negative, which corresponds to the p-type quasiparticle conductivity, which is known to be inherent in TaS_3 [12].

3. EXPERIMENT

Blue bronze crystals differ from other quasi-one-dimensional conductors with CDWs in their relatively large sizes, which facilitates the preparation of point contacts, including those for passage of current through chains. We selected crystals with a perfect surface perpendicular to the b axis of the crystal. In some cases, the surface was obtained by directional cleavage of the initial crystal. Electrochemically sharpened copper wires 50 and 75 μm in diameter were used as normal electrodes. The radius of the tip curvature did not exceed 1 μm. Point contacts were formed directly at a low temperature with the help of a precision mechanical transmission system. The contact diameter d was estimated from the value of its resistance R using the Sharvin formula [13]

$$R \approx \frac{\rho l}{a^2}, \quad a = d\sqrt{A},$$

where $\rho = 0.2\ \Omega$ cm is the resistivity of $K_{0.3}MoO_3$ along the chains at T = 77 K and l = 10 nm is the mean free path [1]. The entire system was in a gaseous heat-exchange medium, in which the temperature can be varied from 77 to 300 K. We prepared and studied contacts in a wide range of contact resistances from units to several hundreds kiloohms at T = 78 K. Applying a voltage to high-resistance contacts ($R \gtrsim$ 30–40 kΩ, $d \lesssim 100$ Å), we could not attain the critical CDW deformation corresponding to the beginning of phase slip. In the case of large-area contacts, we could observe a local CDW sliding in the contact region. It was difficult to obtain temperature dependences for high-resistance contacts due to their low stability (especially upon a change in temperature). In some cases, however, we could trace the characteristics of such contacts in wide temperature ranges. In addition, temperature evolution of the contact characteristics was also studied as follows: the characteristics of two or three contacts were measured at each preset temperature.

4. HIGH-RESISTANCE CONTACTS

It was shown in [8] that an appreciable CDW deformation can be attained for point contacts with a small characteristic size (as a rule, $d \lesssim 100$ Å); in this case, the chemical potential may be shifted below the effective middle of the Peierls gap ($\delta\zeta > \zeta_0$). The measurements were made at the liquid nitrogen temperature. In this section, we describe the results of analysis of characteristics for such contacts upon a change in temperature. Figure 2 shows the $R(V)$ dependences for a contact, for which the measurements could be made in a temperature range of 80–190 K. At a low temperature, the curves are asymmetric and the resistance peak is displaced towards positive voltages.

It can be seen that the voltage V_0 corresponding to the resistance peak decreases upon an increase in temperature and becomes close to zero at $T \approx 140$ K. At $T > 140$ K, the $R(V)$ dependence becomes almost symmetric. It is interesting to note that the nonlinearity of the current–voltage characteristic (IV curve) is also preserved above the Peierls transition temperature T_P =

183 K, which indicates the existence of a fluctuation CDW in $K_{0.3}MoO_3$ at $T > T_P$. In this case, the observed nonlinearity may be associated with a lowered density of states near the Fermi level, which is a result of fluctuations of the CDW order parameter at $T > T_P$ (pseudogap) [14]. In accordance with the results obtained in [15], the fluctuation threshold contribution to conductivity was observed for thin TaS_3 samples at temperatures exceeding the Peierls transition temperature by 40 K.

At $T < 80$–90 K, the $R(V)$ curves become noticeably asymmetric not only about $V = 0$, but also relative to V_0 (i.e., these curves cannot be described by expression (1) any longer). Such a form of the $R(V)$ dependence obviously reflects low-temperature anomalies of CDWs [16]. Such effects will not be considered here. The results of investigation of point contacts between a normal metal and blue bronze at low temperatures are described in [17].

To verify whether or not the observed characteristics of point contacts can be attributed to the properties of microscopic volumes of a Peierls conductor, it is expedient to compare the temperature dependences of the point contact resistance with an analogous dependence obtained for a bulk sample. Light circles in Fig. 3 show the resistance obtained for the point contact whose characteristics are represented in Fig. 2 at $V = 0$, while dark circles correspond to $V = V_0$. The solid straight line corresponds to the activation dependence at an activation temperature of $T_0 = 720$ K, which is in good agreement with the available results of measurement of the Peierls gap $\Delta \approx 50$–70 meV (500–700 K) in $K_{0.3}MoO_3$, which were obtained in tunnel experiments [18] in which the reflection and absorption of light [19] were measured. The obtained value of T_0 is also in good agreement with the activation energy of conduction for bulk blue bronze samples [1, 10]. The experimental values of R for $V = V_0$ at low temperatures are in better agreement with the activation dependence than the values of resistance at $V = 0$. This result could be expected: in accordance with relation (1), the value of $R(V_0)$ corresponds to the position of the chemical potential at the middle of the gap.

To obtain more reliable information in a temperature range of $T = 80$–200 K, we measured the characteristics of various contacts at several preset values of temperature. In spite of the spread in these values, the qualitative pattern of temperature variation of the IV curves was the same for all contacts and corresponded to the data depicted in Fig. 2. For example, the value of voltage V_0 decreased with increasing temperature. At high temperatures ($T > 140$ K), the $R(V)$ dependences for all contacts became almost symmetric relative to the change in the polarity of the applied voltage.

Figure 4 shows by way of example the results obtained for three different contacts at $T = 119.6$ K. At a fixed temperature, the voltage corresponding of the resistance peak on the $R(V)$ curves was almost the same for all contacts and corresponded to approximately the same value of bias voltage.

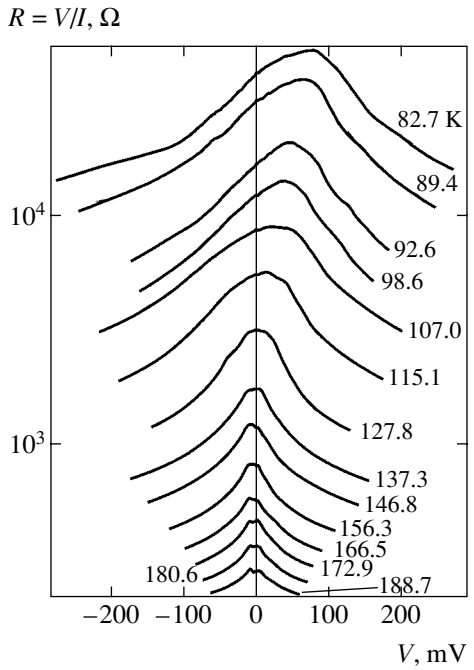


Fig. 2. $R(V)$ dependences for a contact in the temperature range 80–190 K.

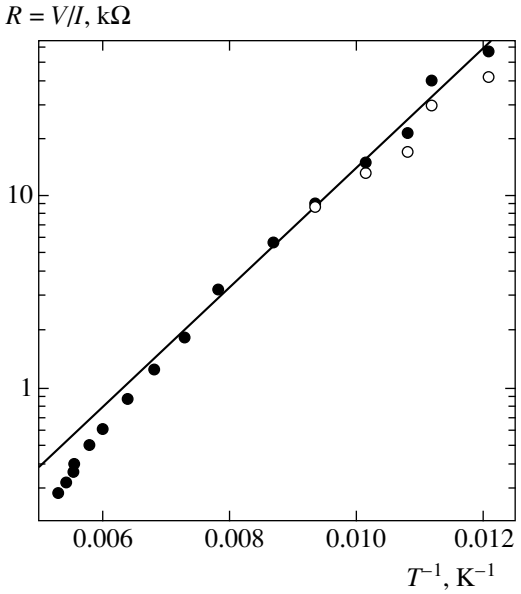


Fig. 3. Temperature dependences of the resistance of the same contact as in Fig. 2 at $V = 0$ (light circles) and $V = V_0$ (dark circles). The slope of the straight line corresponds to the activation energy $T_0 = 720$ K.

5. LOW-RESISTANCE CONTACTS

Point contacts with a resistance $R(V) < 30$ kΩ at nitrogen temperature exhibited a considerably higher

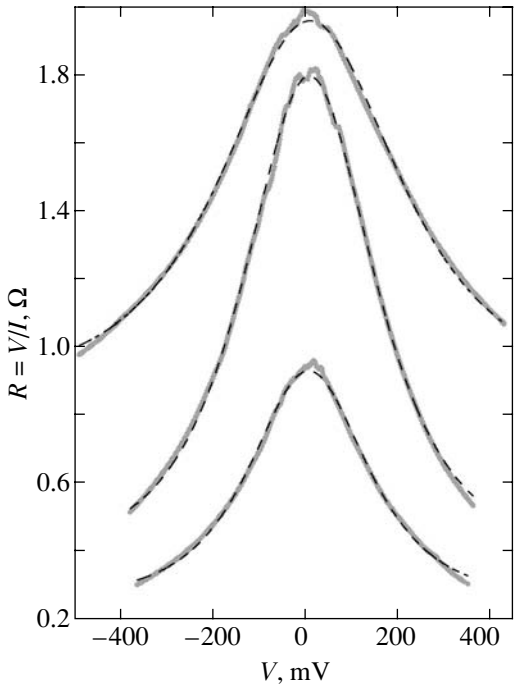


Fig. 4. Experimental $R(V)$ dependences for three different $Cu–K_{0.3}MoO_3$ contacts at $T = 119.6$ K (solid curves). The dashed curves describe the results of fitting with the help of formula (1).

stability, which allowed us to analyze their characteristics in detail in a wide temperature range. A distinguishing feature of the characteristics of such contacts is a sharp decrease in the resistance at a certain bias voltage $\pm V_t$, which is often of the "breakdown" type. For better visualization, it is expedient to plot the differential and not the ordinary resistance of such contacts as a function of voltage. Figure 5 shows the dependence of the differential resistance $R_d = dV/dI$ on the bias voltage at $T = 77$ K for a contact in which this effect is manifested most clearly. For $V = V_t > 0$, the resistance of this contact abruptly decreases by almost an order of magnitude. For $V = -|V_t|$, a sharp variation of R_d can also be seen. However, the amplitude of the resistance jump for the negative bias voltage is much smaller than for the positive bias voltage. In most cases, the jump in differential resistance can be clearly traced for both voltage polarities.

Figure 6 shows the temperature evolution of the $R_d(V)$ dependence for a contact of this type in a temperature range of 80–154 K. It should be noted that the $R_d(T)$ curve for zero bias voltage follows the activation dependence with an activation energy of $T_0 \approx 700$ K in this case also.

At a high temperature ($T > 130$ K), the dependences become almost symmetric about $V = 0$ and the curves are qualitatively similar to differential IV curves for a typical conductor with CDWs. In the latter case, a sharp decrease in the differential resistance is observed after the attainment of the threshold electric voltage corresponding to the beginning of CDW slip. Figure 7 shows the temperature dependence of voltage V_t for the same contact.

In some cases, we could trace the variation of the parameters of contacts upon a decrease in the resistance

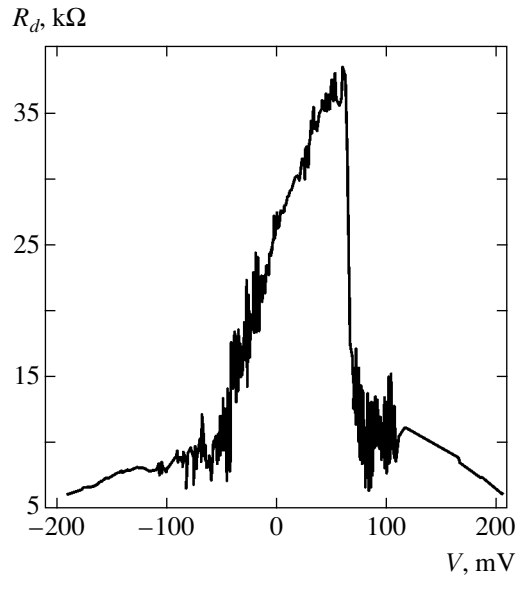


Fig. 5. $R_d(V)$ dependences for a $Cu–K_{0.3}MoO_3$ contact with $R = 26$ kΩ at $T = 77$ K and $V = 0$.

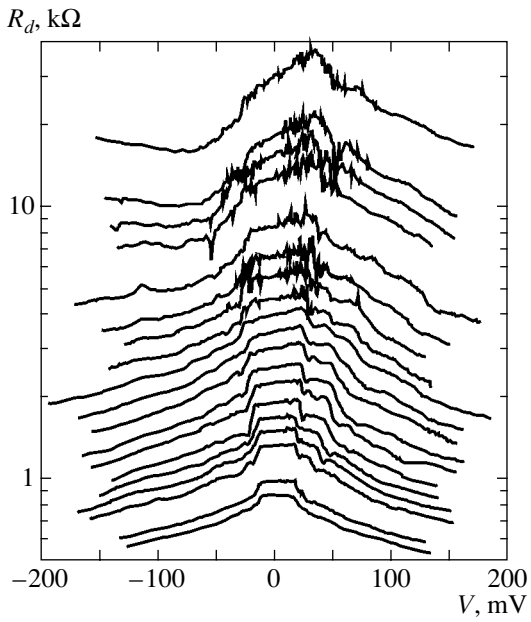


Fig. 6. Temperature evolution of the $R_d(V)$ dependence for a low-resistance $Cu–K_{0.3}MoO_3$ contact. The temperature increases from 80 to 154 K in the downward direction.

(i.e., upon an increase in their area). The latter was attained by smoothly increasing the force applied to the tip of the normal electrode. Figure 8 shows the differential IV curves obtained at $T = 77$ K for one of the contacts. It can be seen that, upon a decrease in the contact resistance, the value of V_t monotonically decreases.

6. DISCUSSION

We begin the discussion of results with high-resistance contacts. Using formula (1), we obtained the theoretical $R(V)$ dependences and varied the fitting parameters to attain the best agreement with the experimental data. Dashed curves in Fig. 4 illustrate the results of this procedure. Dark circles in Fig. 9 show the initial position of the chemical potential at various temperatures, obtained as a result of the above procedure for the sample whose $R(V)$ dependences are depicted in Fig. 2. Light circles in the same figure represent the data obtained as a result of fitting for the $R(V)$ dependences measured at these temperatures for various contacts. The spread in the values in the latter case can be explained by the fact that pinning leads to spatially nonuniform deformation of CDWs, which can be manifested in spatial fluctuations of the chemical potential on the order of $\sqrt{E_t(d\zeta/dq)}$ [20], which may reach 1 meV at low temperatures. However, qualitative agreement between the results obtained in both types of experiments can be clearly seen. In further analysis of the results, we will disregard the initial deformation (at zero bias voltage).

It can be seen from Fig. 9 that the upward shift in the chemical potential decreases upon heating and virtually vanishes at $T \approx 140$ K. Above this temperature, negative values of V_0 were obtained in some cases (the initial position of the chemical potential is lower than the middle of the energy gap). A certain additional shift of the chemical potential may be due to the initial CDW deformation (at zero bias voltage) discussed above.

However, the most probable explanation is as follows. With increasing temperature, the resistance of the contact region of blue bronze decreases. The relative contribution from the resistance of the boundary increases accordingly. In this case, one should expect a manifestation of barrier (tunnel) effects in the characteristics of point contacts (namely, the effect of reflection of normal carriers injected from the normal metal from the barrier associated with the Peierls energy gap) [21, 22]. The reflection of carriers is manifested in the experiment in the form of excess differential resistance of the contact between the normal metal and the semiconductor with CDWs in the region of $|V| < \Delta$. This effect was reliably observed in the system with CDWs with incomplete dielectrization of the electron spectrum (in $NbSe_3$) [23, 24], in which the presence of non-

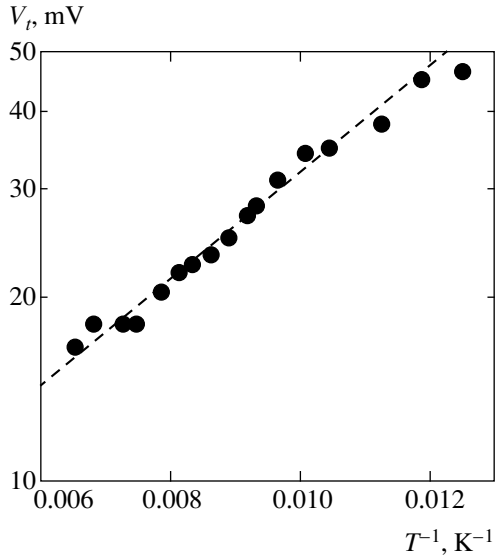


Fig. 7. Temperature dependence of the bias voltage V_t applied to the contact whose IV curves are shown in Fig. 6. The dashed line describes the function $V = V_0\exp(T_0/T)$, where $T_0 = 200$ K.

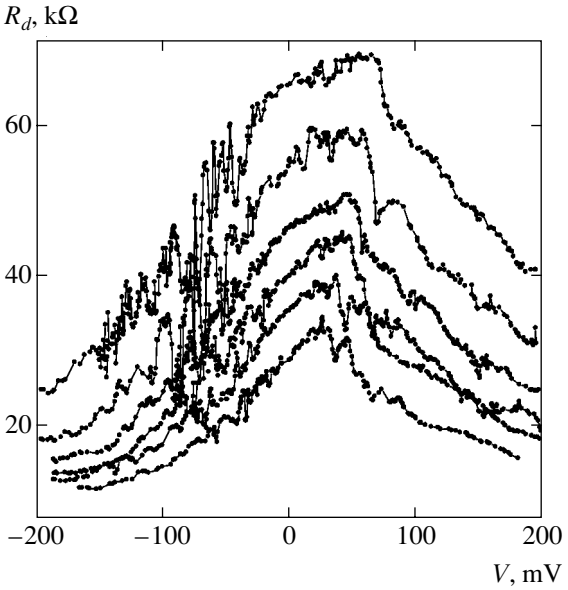


Fig. 8. $R_d(V)$ dependences for a Cu–$K_{0.3}MoO_3$ contact upon a decrease in its resistance (increase in diameter) at $T = 77$ K.

condensed carriers in CDWs ensures effective screening of the electric field and, hence, the absence of a noticeable band distortion in the contact region.

Figure 10 shows the curves describing the differential resistance of the contact whose characteristics are depicted in Fig. 2 in the range of low bias voltages at temperatures close to T_P. It can be seen that, as we approach T_P, the curves become more and more asymmetric relative to the R_d peak and exhibit a fine struc-

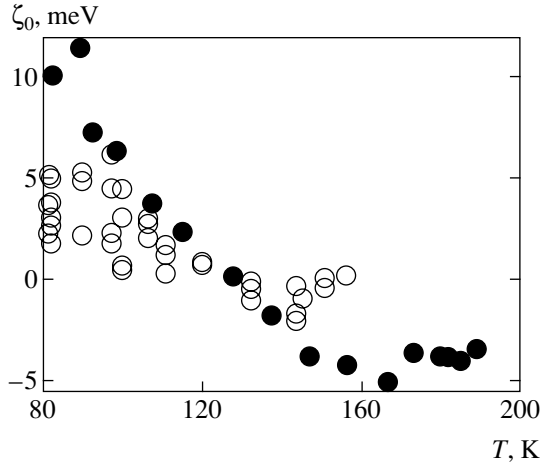


Fig. 9. Temperature dependence of the initial chemical potential shift relative to the middle of the energy gap for the contact shown in Fig. 2 (dark circles) and for various contacts formed at fixed temperatures (light circles).

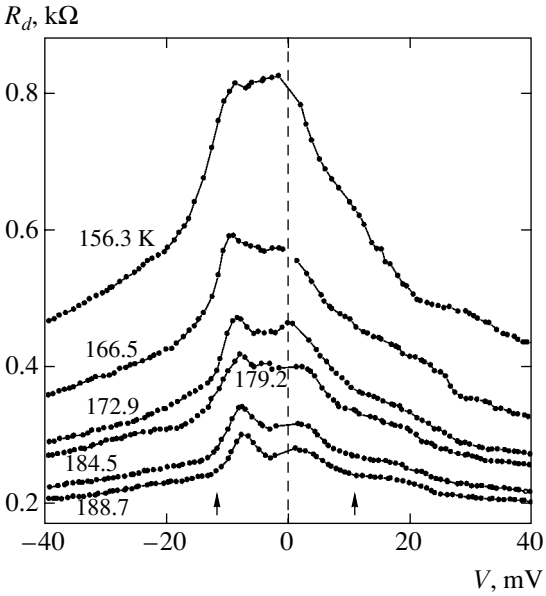


Fig. 10. Temperature dependence of the differential IV curves of the contact shown in Fig. 2 in the region of small bias voltages at a temperature close to T_P. The arrows indicate approximate positions of voltages corresponding to the energy gap.

ture. An asymmetric form of the IV curves for a tunnel structure of the metal–insulator–CDW type was recently obtained theoretically in [25]; for $|V| > \Delta/e$, where oscillations of the density of states are predicted (!). Such a behavior can be explained by the presence of the open CDW boundary.

Differential IV curves of a direct metal–CDW contact (without an insulating barrier) also reflect peculiarities of the density of states, which is confirmed by the results obtained by studying materials with CDWs and with a semimetallic ground state [23, 24]. In the case of complete dielectrization of the spectrum upon the Peierls transition (the case with $K_{0.3}MoO_3$) at low temperature, these features are indistinguishable against the background of much stronger effects of CDW deformation in the contact region. Tunnel effects in the metal–$K_{0.3}MoO_3$ point contact can be manifested at temperatures close to T_P. Obviously, the energy gap is strongly blurred at so high temperatures. Arrows in Fig. 10 denote the approximate positions of the bias voltages corresponding to the Peierls energy gap. It can be seen that this singularity in IV curves is manifested at $T \approx 160$ K. It should be noted once again that these features of IV curves corresponding to the Peierls energy gap are also observed at $T > T_P$, which may indicate the existence of the fluctuation order parameter of CDWs.

As a result of manifestation of tunnel effects, the dependences of resistance $R = V/I$ on the bias voltage in a temperature range near T_P are also asymmetric with the resistance peak displaced to the region of negative voltages. Consequently, the description of characteristics of contacts only in the framework of the semiconductor model based on formula (1) disregarding tunnel effects in the temperature range 160–190 K is not quite correct.

An interesting result was obtained for the screening coefficient b in approximating the $R(V)$ dependences by formula (1). Figure 11 shows the temperature dependence of coefficient b for the same contact as in Fig. 2. It can be seen that screening of the field by a CDW decreases with temperature, which may indicate an increase in transverse rigidity and, hence, coherence of the CDW. Indeed, the electric field deforms the CDW in the local region in the vicinity of the contact. However, owing to the interaction with neighboring chains (pinning), the deformation region turns out to be larger and increases upon cooling. This process can be treated as an effective increase in the coherence volume.

Conversely, CDW screening at $T > 140$ K becomes almost complete; i.e., the electrostatic potential does not change at the sample surface. The screening parameters obtained for individual contacts at fixed temperatures exhibited a considerable spread in the values of b and are not shown in the figure. However, the general tendency towards a decrease in screening (coefficient b) upon cooling could be traced in all cases.

The electric field penetration depth in the semiconductor with CDWs increases with the contact diameter. Phase slip is facilitated in this case. Indeed, a single event of phase slip corresponds to the emergence or disappearance of a CDW period. In this case, the change in the wavevector can be estimated as $\delta q \approx 2\pi/a$, where $a \approx d\sqrt{A}$ is the diameter of the region of field penetration to the bulk of the sample; the corresponding value

of $\delta\zeta_{2\pi}$ can easily be determined if we know the temperature dependence $R(T)$ of the resistance:

$$\delta\zeta_{2\pi} \approx \delta q \frac{d\zeta}{dq} \approx kT \frac{2\pi}{aq} \frac{R(T)}{R(300\ \mathrm{K})}. \quad (2)$$

The value for $d\zeta/dq$ was borrowed from [20], where the temperature dependence of mobilities μ was ignored. Until $\delta\zeta$ reaches a value approximately equal to $\delta\zeta_{2\pi}$, phase slip, as well as the propagation of the CDW, is ruled out. The value of $\delta\zeta_{2\pi}$ is inversely proportional to a and, hence, to the contact diameter d. As the value of d increases, it becomes possible to attain the critical deformation of the CDW at a certain bias voltage V_t. In a certain sense, this voltage can be treated as the critical voltage for the onset of local (over a length on the order of a) CDW sliding. At a lower voltage (i.e., smaller value of $\delta\zeta$), the appearance or disappearance of a CDW period over length a is disadvantageous since it would lead to the emergence of an even stronger CDW deformation of the opposite sign.

The above scenario of CDW depinning determined by phase slip is apparently realized in low-resistance contacts whose characteristics are depicted in Figs. 5, 6, and 8. Estimate (2) gives a reasonable value of $\delta\zeta_{2\pi}$; substituting T = 120 K, $R(T)/R(300\ \mathrm{K})$ = 30, and a = 300 Å into this formula, we obtain $\delta\zeta_{2\pi} \approx 10$ meV, which is comparable with the value of $V_t b$ (see Figs. 7 and 11). To obtain this estimate, the value of the bulk threshold field is not required; i.e., CDW breakdown is not determined by impurity pinning in the given model.

The temperature behavior of the threshold voltage V_t in point contacts differs from the temperature evolution of the threshold field E_t in bulk $K_{0.3}MoO_3$ samples. Indeed, the value of E_t for blue bronze at $T > 50$ K either increases monotonically with temperature [10], or increases in the interval 50 K < T < 100 K and decreases upon a further increase in temperature [26] depending on the crystal quality. It can be seen in Fig. 7 that the value of V_t in our case monotonically decreases in the entire temperature range 80–160 K. This is not astonishing since the beginning of CDW slip in the case of a point contact is determined by phase slip rather than by CDW pinning. Expression (2) qualitatively describes the increase in V_t upon cooling. From the standpoint of physics, such a behavior is associated with an increase in the CDW elastic modulus $d\zeta/dq$. The experimental results are satisfactorily described by a dependence of the type $V = V_0 \exp(T_0/T)$, where T_0 = 200 K (dashed straight line in Fig. 7). This activation energy is less than half the value of Δ, which contradicts formula (2). Some experimental data indicate, however, that the CDW elastic modulus increases upon a decrease in temperature at a lower rate [27]. The decrease in the value of $d\zeta/dq$ is probably associated with a finite density of states in the gap. The existence of such states in semiconductors with CDWs was predicted theoretically in [28] for amplitude solitons. In addition, fluctuation-induced blurring of the edges of the electron and hole bands should also be taken into account.

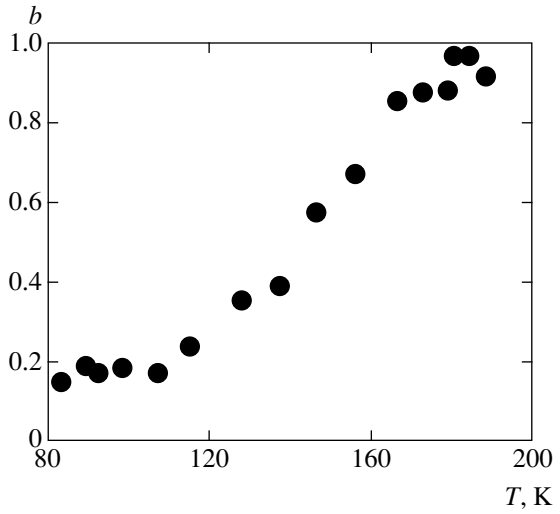


Fig. 11. Temperature dependence of the screening coefficient b for the contact whose characteristic are given in Fig. 2.

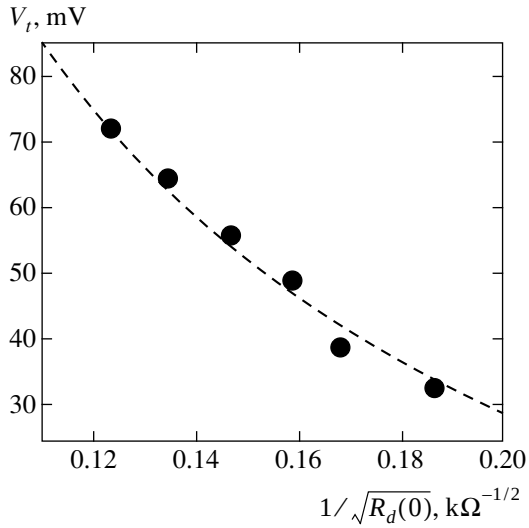


Fig. 12. Dependence of the threshold voltage V_t on $1/\sqrt{R_d(0)} \propto d$ for the contact shown in Fig. 8 at T = 77 K. The dashed curve corresponds to the hyperbolic dependence $V_t \propto 1/d$.

Figure 12 shows the dependence of voltage V_t on the quantity $1/\sqrt{R_d(0)}$ proportional to the contact diameter [13]. It can be seen that the threshold voltage monotonically decreases with increasing contact diameter, which also agrees with expression (2). A increase in the threshold field upon a decrease in the transverse size of the samples was observed earlier in the study of the size effect in $NbSe_3$ [29–31] and in TaS_3 [4]; it was concluded that $E_t \propto 1/d$, where d is the sample thickness. It

can be seen from Fig. 12 that our experimental values of V_t are also in good agreement with the hyperbolic dependence (dashed curve). However, in accordance with relation (2), CDW breakdown is associated not with overcoming of bulk pinning, but with the size limitation imposed on phase slip (i.e., phase slip occurs over a limited length). Thus, relation (2) describes the dependences of V_t on the temperature and size of the contact; these dependences can be explained by the size effect in phase slip.

7. CONCLUSIONS

Thus, we have studied and analyzed the evolution of the properties of metal–CDW point contacts upon a change in the contact diameter and temperature. In the temperature range 80–150 K, the experimental data can be adequately described in the framework of the semiconductor model of CDWs [5]. We demonstrated that the contact resistance is determined by a microscopic region whose size is determined by the contact diameter and anisotropy in the conductivity of a Peierls sample. It is in this region that the conductivity changes appreciably under the action of the electric field: screening of the external electric field leads to CDW deformation, chemical potential shift for quasiparticles, and a change in the point contact resistance. The temperature dependence of the position of the chemical potential has been determined. It is shown that the chemical potential in equilibrium is above the middle of the Peierls gap and approaches its middle upon an increase in temperature. It is found that local movement of CDWs in the contact region can be observed only for large values of the contact diameter ($d \gtrsim 100$ Å). For this type of contacts, we have determined the dependence of the CDW breakdown voltage on the contact diameter and temperature and proposed a model explaining these dependences by the size effect in the CDW phase slip. At temperatures close to T_P, the differential IV curves of the contacts reveal the features of the spectrum of single-particle excitations. The observed nonlinearity of the IV curves at $T > T_P$ is explained by the existence of the CDW fluctuation order parameter.

On the whole, the results open a new field in CDW physics, i.e., semiconductor and dynamic properties of CDWs in microscopic volumes under huge electric fields and CDW deformation. It is shown that a point contact can be used as a local probe for studying the energy structure of CDWs. Point-contact studies of CDWs are being continued. It would be interesting to study point contacts using metals with various values of the work function as well as samples with various values of the bulk threshold field E_t as normal counterelectrodes. Of special importance is the study of metal–CDW point contacts at temperatures below liquid nitrogen temperature.

ACKNOWLEDGMENTS

The authors are grateful to R.E. Thorne for providing high-quality blue bronze samples and to S.A. Brazovskiĭ, A.V. Golovnya, S.V. Zaitsev–Zotov, and S.G. Zybtsev for fruitful discussion of the results.

This study was supported by the Russian Foundation for Basic Research (project nos. 05-02-17578 and 03-02-22001 CNRS), INTAS (grant no. 01-0474), and the programs "New Materials and Structures" of the Presidium of the Russian Academy of Sciences and Universities of Russia (grant no. 146-05).

REFERENCES

1. *Charge Density Waves in Solids*, Ed. by L. Gor'kov and G. Grüner (Elsevier, Amsterdam, 1989); G. Grüner, *Density Waves in Solids* (Addison-Wesley, Reading, Mass., 1994).
2. S. N. Artemenko and A. F. Volkov, Zh. Éksp. Teor. Fiz. **81**, 1872 (1981) [Sov. Phys. JETP **54**, 992 (1981)].
3. M. E. Itkis, F. Ya. Nad', and V. Ya. Pokrovskiĭ, Zh. Éksp. Teor. Fiz. **90**, 307 (1986) [Sov. Phys. JETP **63**, 177 (1986)].
4. D. V. Borodin, S. V. Zaĭtsev-Zotov, and F. Ya. Nad', Zh. Éksp. Teor. Fiz. **93**, 1394 (1987) [Sov. Phys. JETP **66**, 793 (1987)].
5. S. N. Artemenko, S. V. Zaĭtsev-Zotov, and V. Ya. Pokrovskiĭ, Zh. Éksp. Teor. Fiz. **110**, 1069 (1996) [JETP **83**, 590 (1996)].
6. H. Requardt, F. Ya. Nad, P. Monceau, *et al.*, Phys. Rev. Lett. **80**, 5631 (1998).
7. S. Brazovskii, N. Kirova, H. Requardt, *et al.*, Phys. Rev. B **61**, 10640 (2000).
8. A. A. Sinchenko, V. Ya. Pokrovskiĭ, and S. G. Zybtsev, Pis'ma Zh. Éksp. Teor. Fiz. **74**, 191 (2001) [JETP Lett. **74**, 173 (2001)].
9. I. O. Kulik, A. N. Omel'yanchuk, and R. I. Shekhter, Fiz. Nizk. Temp. **3**, 1543 (1977) [Sov. J. Low Temp. Phys. **3**, 740 (1977)].
10. C. Schlenker, in *Low-Dimensional Electronic Properties of Molybdenum Bronzes and Oxides* (Kluwer Academic, Dordrecht, 1989).
11. A. A. Sinchenko, Yu. I. Latyshev, V. Ya. Pokrovskii, *et al.*, J. Phys. A **36**, 9311 (2003).
12. A. W. Higgs, in *Springer Lecture Notes in Physics*, Ed. by G. Hutirai and J. Solyom (1985), Vol. 217, p. 422; Yu. I. Latyshev, Ya. S. Savitskaya, and V. V. Frolov, Pis'ma Zh. Éksp. Teor. Fiz. **38**, 446 (1983) [JETP Lett. **38**, 541 (1983)].
13. Yu. V. Sharvin, Zh. Éksp. Teor. Fiz. **48**, 984 (1965) [JETP Lett. **21**, 655 (1965)].
14. B. P. Gorshunov, A. A. Volkov, G. V. Kozlov, *et al.*, Phys. Rev. Lett. **73**, 308 (1994).
15. V. Ya. Pokrovskii, S. V. Zaitzev-Zotov, and P. Monceau, Phys. Rev. B **55**, R13377 (1997).
16. S. V. Zaitsev-Zotov, Pis'ma Zh. Éksp. Teor. Fiz. **46**, 453 (1987) [JETP Lett. **46**, 572 (1987)]; Phys. Rev. Lett. **71**, 605 (1993); N. Ogawa, A. Shiraga, R. Kondo, *et al.*,

Phys. Rev. Lett. **87**, 256401 (2001); J. Odin, J. C. Lasjaunias, K. Biljaković, *et al.*, Eur. Phys. J. **24**, 315 (2001); D. Starešinić, K. Biljaković, W. Brütting, *et al.*, Phys. Rev. B **65**, 165109 (2002).

17. A. A. Sinchenko and V. Ya. Pokrovskii, in *Abstracts of the International Workshop on Electronic Crystals ECRYS-2005, Corsica, France, 2005*, p. 49; J. Phys. IV (in press).
18. A. A. Sinchenko, Yu. I. Latyshev, and S. G. Zybtsev, Zh. Éksp. Teor. Fiz. **113**, 1830 (1998) [JETP **86**, 1001 (1998)].
19. G. Travaglini, P. Wachter, P. Marcus, and C. Shlenker, Solid State Commun. **37**, 599 (1981).
20. V. Ya. Pokrovskii and S. V. Zaitsev-Zotov, Synth. Met. **32**, 321 (1989).
21. M. I. Visscher and G. E. W. Bauer, Phys. Rev. B **54**, 2798 (1996).
22. B. Rejaei and G. E. W. Bauer, Phys. Rev. B **54**, 8487 (1996).
23. A. A. Sinchenko, Yu. I. Latyshev, S. G. Zybtsev, *et al.*, Phys. Rev. B **60**, 4624 (1999).
24. A. A. Sinchenko and P. Monceau, Phys. Rev. B **67**, 125117 (2003).
25. B. Dora, cond-mat/0503435.
26. J. Dumas, C. Schlenker, J. Marcus, and R. Buder, Phys. Rev. Lett. **50**, 757 (1983).
27. B. Hennion, J. P. Pouget, and M. Sato, Phys. Rev. Lett. **68**, 2374 (1992); A. Virosztek and K. Maki, Phys. Rev. Lett. **69**, 3265 (1992); A. V. Golovnya, V. Ya. Pokrovskii, and P. M. Shadrin, Phys. Rev. Lett. **88**, 246401 (2002).
28. S. I. Matveenko and S. A. Brazovskii, Phys. Rev. B **65**, 245108 (2002).
29. J. McCarten, D. A. DiCarlo, M. P. Maher, *et al.*, Phys. Rev. B **46**, 4456 (1992).
30. P. J. Yetman and J. C. Gill, Solid State Commun. **62**, 201 (1987).
31. R. E. Thorne and J. McCarten, Phys. Rev. Lett. **65**, 272 (1990).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 101, No. 6, 2005, pp. 1140–1146.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 6, 2005, pp. 1299–1306.

STATISTICAL, NONLINEAR, AND SOFT MATTER PHYSICS

Nonlinear Dynamics of Convection Patterns in a Binary Mixture Subjected to Finite-Frequency Vibration

B. I. Myznikova[a] and B. L. Smorodin[b]

[a]*Institute of Continuous Media Mechanics, Ural Division, Russian Academy of Sciences, Perm, 614013 Russia*

[b]*Perm State University, Perm, 614990 Russia*

e-mail: smorodin@psu.ru

Received June 16, 2005

Abstract—Nonlinear evolution of two-dimensional convection patterns is considered for an incompressible binary mixture with negative Soret coupling in a horizontal layer subjected to finite-frequency vertical vibration of arbitrary amplitude. A numerical analysis is performed under impermeability conditions on rigid boundaries, which can be implemented in a laboratory experiment. The dependence of flow intensity on vibration amplitude is examined for the first and second resonance regions in the parameter space of thermal vibrational convection. The numerical results agree with the stability boundaries of equilibrium states predicted by linear theory. A qualitative difference in the dynamics of nonlinear oscillation is exposed between the regions corresponding to critical perturbations at the subharmonic and fundamental frequencies. Regular and chaotic dynamics, as well as hysteretic transitions between the fundamental and subharmonic modes, are revealed. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Investigations of phenomena leading to pattern formation in homogeneous systems date back to studies by Bénard and von Kármán. Self-organization processes in multicomponent fluid mixtures lead to various flow regimes. Recent experimental [1–3] and theoretical [4–7] investigations of convection in binary fluid mixtures have revealed regimes characterized by diverse spatial and temporal behavior. The complexity of convection patterns has been attributed to thermodiffusion (Soret effect), which manifests itself as a concentration gradient induced by a temperature gradient. In the case of negative Soret coupling, when diffusion of the lighter component toward the "cold" boundary reduces buoyancy effects, binary-mixture convection under constant gravity can be caused by oscillatory instability. Nonlinear evolution of perturbations under supercritical conditions can result in the formation of stationary patterns, standing or traveling waves, localized traveling waves, and "chevrons."

In applications, the performance of technological systems can be altered by vibration, ac electric field, or temperature modulation. The Kapitza pendulum is a classical example of vibration-induced qualitative change in the response of a physical system to perturbation of its state of equilibrium [8]. Stability and nonlinear dynamics of convective systems are also strongly modified by external modulation [9], which can be used to control fluid dynamics and heat or mass transfer in various technological systems.

The effect of vibration on the regime of binary fluid convection, steady flow pattern, and the possibility of its destabilization is a subject of considerable interest [10]. The analysis presented in [11] was focused on convective stability of a quasi-equilibrium state and supercritical regimes of Soret-driven binary-mixture convection between impermeable horizontal solid plates in the limit of high-frequency vibration with interdependent amplitude and frequency under off-resonance conditions. It was shown that thermally driven convection in a binary mixture with negative Soret coupling evolves into a stationary pattern rather than an oscillatory regime of thermal vibrational convection. The threshold for the onset of this pattern differs from the critical condition predicted by linear stability theory and depends on the nonlinear dynamics of the system.

In this paper, we analyze the nonlinear stage of convection driven by a temperature gradient across a horizontal layer of a mixture with negative Soret coupling between impermeable solid boundaries. The system is subjected to finite-frequency vertical vibration of arbitrary amplitude. We examine the effect of vibration amplitude on the evolution of convection patterns corresponding to different resonance regions in the parameter space and reveal hysteretic transitions between nonlinear oscillations at the fundamental and subharmonic frequencies, as well as transitions between regular and chaotic dynamics.

1063-7761/05/10106-1140$26.00

2. STATEMENT OF THE PROBLEM

Consider the layer of a binary fluid mixture between perfectly heat-conducting impermeable solid plane boundaries located at $z = 0$ and $z = h$ and held at constant temperatures $T(0) = \Theta$ and $T(h) = 0$. (The cases of $\Theta < 0$ and $\Theta > 0$ correspond to heating from above and below, respectively.) Even though no concentration gradient is imposed, the relative motion of the mixture components due to thermodiffusion (Soret effect) gives rise to a concentration gradient even in an initially homogeneous mixture [12].

The equation of state is

$$\rho = \bar{\rho}(1 - \beta_T T - \beta_C C), \tag{1}$$

where $\bar{\rho}$ is the mixture density corresponding to the mean values of temperature and concentration; T and C are the deviations of temperature and concentration from their respective mean values; and

$$\beta_T = -\left(\frac{1}{\bar{\rho}}\frac{\partial \rho}{\partial T}\right), \quad \beta_C = -\left(\frac{1}{\bar{\rho}}\frac{\partial \rho}{\partial C}\right)$$

are the thermal and solutal expansion coefficients, respectively. Assuming that C is the concentration of the lighter component, we have $\beta_C > 0$.

The layer harmonically oscillates along the z axis, with frequency Ω and amplitude b. In the oscillating frame of reference, the modulated gravitational acceleration is $\mathbf{g} + b\Omega^2 \sin(\Omega t)\mathbf{n}$, where $\mathbf{g}$ is the static gravitational acceleration and $\mathbf{n}$ is the unit vector along the z axis.

The equation of motion for the binary mixture contains terms representing the forces due to concentration gradients in a modulated gravity field. In addition to the heat and continuity equations, the model includes an evolution equation for the concentration of the lighter component in which the concentration flux contains both Fickian and Soret diffusion terms:

$$\mathbf{j} = -\bar{\rho} D(\nabla C + \alpha \nabla T),$$

where D is the diffusion coefficient and α is the Soret coefficient. The temperature and concentration gradients are antiparallel and parallel in systems with $\alpha < 0$ and $\alpha > 0$ (positive and negative Soret coupling), respectively. In what follows, both D and α are assumed to be independent of temperature.

Using h, h^2/ν, χ/h, Θ, $\beta_T\Theta/\beta_C$, and $\bar{\rho}\nu\chi/h^2$ (where ν is kinematic viscosity and χ is thermal diffusivity) as reference length, time, velocity, temperature, and pressure, respectively, we write the convection equations in the Boussinesq approximation as

$$\frac{\partial \mathbf{v}}{\partial t} + \frac{1}{\mathrm{Pr}}(\mathbf{v}\cdot\nabla)\mathbf{v} = -\nabla p + \Delta \mathbf{v} + (\mathrm{Gr} + \mathrm{Gr}_v \sin\omega t)(T + C)\mathbf{n},$$

$$\frac{\partial T}{\partial t} + \mathbf{v}\cdot\nabla T = \frac{1}{\mathrm{Pr}}\Delta T, \quad \mathrm{div}\mathbf{v} = 0,$$
$$\frac{\partial C}{\partial t} + \mathbf{v}\cdot\nabla C = \frac{1}{\mathrm{Sc}}\Delta(C - \psi T), \tag{2}$$

$$\mathrm{Gr} = \frac{g\beta\Theta h^3}{\nu^2}, \quad \mathrm{Gr}_v = \frac{b\Omega^2\beta\Theta h^3}{\nu^2}, \quad \mathrm{Pr} = \frac{\nu}{\chi},$$
$$\psi = -\frac{\beta_C \alpha}{\beta_T}, \quad \mathrm{Sc} = \frac{\nu}{D}, \quad \omega = \frac{\Omega h^2}{\nu}.$$

Here, $\mathbf{v}$ is the flow velocity, and p is the dynamic pressure. The dimensionless parameters include the separation ratio ψ ($\psi > 0$ and $\psi < 0$ correspond to positive and negative Soret coupling, respectively), Gr is the Grashof number, Gr_v is the vibrational Grashof number, Pr is the Prandtl number, and Sc is the Schmidt number.

Boundary conditions of different types can be used to analyze binary fluid convection. In terms of experimental feasibility, the most realistic conditions correspond to impermeable isothermal rigid boundaries:

$$z = 0: \quad \mathbf{v} = 0, \quad T = 1, \quad \frac{\partial C}{\partial z} - \psi\frac{\partial T}{\partial z} = 0,$$
$$z = 1: \quad \mathbf{v} = 0, \quad T = 0, \quad \frac{\partial C}{\partial z} - \psi\frac{\partial T}{\partial z} = 0. \tag{3}$$

In a state of hydrostatic equilibrium, the vertical concentration and temperature gradients are balanced:

$$\mathbf{v} = 0, \quad T_0 = -z, \quad C_0 = -\psi(z - 1/2),$$
$$p_0 = \frac{(\mathrm{Gr} + \mathrm{Gr}_v \sin\omega t)z^2}{2} + \mathrm{const}. \tag{4}$$

Pressure is eliminated by introducing the streamfunction Ψ defined by the relations

$$v_x = \frac{\partial \Psi}{\partial z}, \quad v_z = -\frac{\partial \Psi}{\partial x}. \tag{5}$$

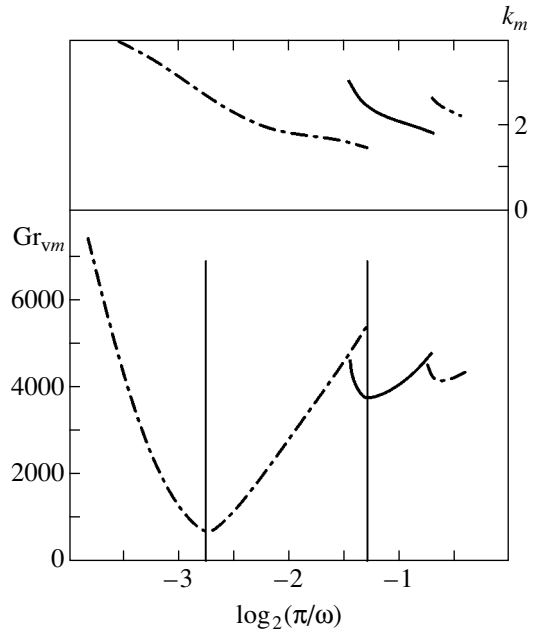


Fig. 1. Critical values of wavenumber and Grashof-number modulation amplitude vs. inverse modulation frequency for $\psi = -0.3$ (negative Soret coupling). In the absence of vibration, the system is in a state of stable equilibrium (Gr = 4000).

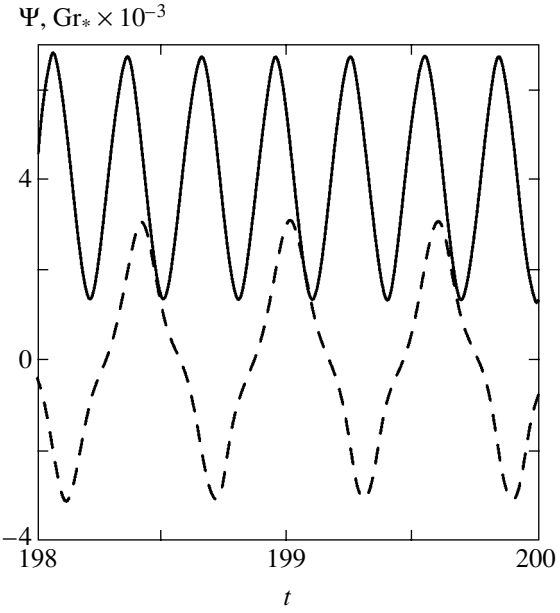


Fig. 2. Time variation of streamfunction (dashed curve) and modulated Grashof number (solid curve) at a point inside the convection cell: $\mathrm{Gr}_v = 2720$, $\omega = 21.10$ (first resonance region).

Then, the system of equations describing thermal vibrational convection in a binary mixture is rewritten in terms of the vorticity y-component $\varphi = (\mathrm{curl}\,\mathbf{v})_y$, temperature, and concentration as follows:

$$\frac{\partial \varphi}{\partial t} + \frac{1}{\mathrm{Pr}}\left(\frac{\partial \Psi}{\partial z}\frac{\partial \varphi}{\partial x} - \frac{\partial \Psi}{\partial x}\frac{\partial \varphi}{\partial z}\right) = \Delta\varphi + (\mathrm{Gr} + \mathrm{Gr}_v \sin\omega t)\frac{\partial (T + C)}{\partial x},$$

$$\frac{\partial T}{\partial t} + \left(\frac{\partial \Psi}{\partial z}\frac{\partial T}{\partial x} - \frac{\partial \Psi}{\partial x}\frac{\partial T}{\partial z}\right) = \frac{1}{\mathrm{Pr}}\Delta T, \tag{6}$$

$$\frac{\partial C}{\partial t} + \left(\frac{\partial \Psi}{\partial z}\frac{\partial C}{\partial x} - \frac{\partial \Psi}{\partial x}\frac{\partial C}{\partial z}\right) = \frac{1}{\mathrm{Sc}}\Delta(C - \psi T),$$

$$\Delta\Psi + \varphi = 0.$$

The boundary conditions are rewritten accordingly:

$$z = 0: \quad \Psi = 0, \quad \frac{\partial \Psi}{\partial z} = 0, \quad T = 1, \quad \frac{\partial C}{\partial z} - \psi\frac{\partial T}{\partial z} = 0,$$

$$z = 1: \quad \Psi = 0, \quad \frac{\partial \Psi}{\partial z} = 0, \quad T = 0, \quad \frac{\partial C}{\partial z} - \psi\frac{\partial T}{\partial z} = 0, \tag{7}$$

$$\Psi(0, z) = \Psi(L, z), \quad \varphi(0, z) = \varphi(L, z),$$
$$T(0, z) = T(L, z), \quad C(0, z) = C(L, z).$$

To compute convection in a cell of length L, we use periodic boundary conditions on the vertical cell boundaries, which facilitates comparison of nonlinear dynamics of convection patterns with the behavior of critical perturbations with wavenumber $k = 2\pi/L$ predicted by the linear analysis of stability of a binary mixture under modulated gravity.

3. NONLINEAR EVOLUTION OF FUNDAMENTAL-FREQUENCY AND SUBHARMONIC PERTURBATIONS

The results presented in this paper were obtained for Pr = 0.75 and Sc = 1.5, which are characteristic of a gaseous mixture.

Figure 1 shows the critical wavenumbers and boundaries of resonance regions in the parameter space corresponding to the onset of instability predicted by the Floquet theory for a mixture with negative Soret coupling ($\psi < 0$).

When $\mathrm{Gr}_v = 0$, the system with $\psi = -0.3$ becomes convectively unstable as the Grashof number increases to $\mathrm{Gr}_m = 4347$. The corresponding neutral perturbations have the frequency $\omega_0 = 10.64$. At $\mathrm{Gr} = 4000 < \mathrm{Gr}_m$, the binary mixture remains at rest. The hydrostatic equilibrium of the mixture becomes unstable with increasing

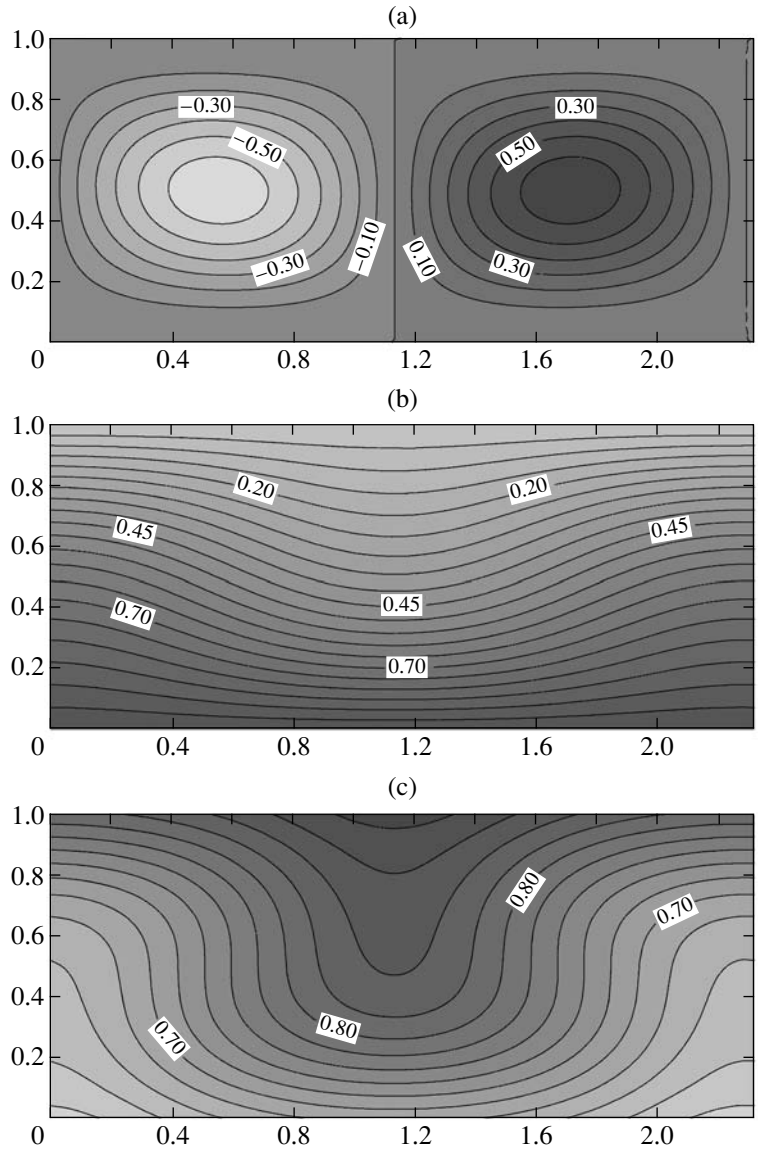


Fig. 3. Visualization of (a) streamfunction, (b) temperature, and (c) concentration fields in the subharmonic mode: $Gr_v = 2720$, $\omega = 21.10$ (first resonance region).

Gr_v. The strongest destabilizing effect of resonant modulation is predicted for $\omega = 21.10 \approx 2\omega_0$ (subharmonic resonance). Under this condition, the horizontal length scale of the convective pattern is determined by the critical wavenumber $k_m = 2.71$, and the corresponding critical vibrational Grashof number $Gr_{vm} = 680.30$ is substantially lower than that in the case of positive Soret coupling. The minimum of the neutral curve corresponding to the fundamental resonance region is located at $\omega = 7.75 \approx \omega_0$, and the corresponding critical parameters are $k_m = 2.48$ and $Gr_{vm} = 3754$. Additional minima correspond to $2\omega_0/m$, where m is an integer. In those regions, the destabilizing effect is weaker. The critical wavenumbers for parametric instability with respect to subharmonic and fundamental-frequency perturbations shown in the top panel of Fig. 1 monotonically decrease with increasing perturbation period within the respective instability regions.

To compute the vorticity, temperature, and concentration fields, we solved nonlinear problem (6), (7) by the fractional time-stepping method, using an alternating-direction implicit scheme. The streamfunction was found by the successive overrelaxation method at each time step.

The parameters used in computer simulations of finite-amplitude convection regimes were taken from the results of the linear analysis. Numerical simulations were performed for parameter values lying on the vertical lines in Fig. 1.

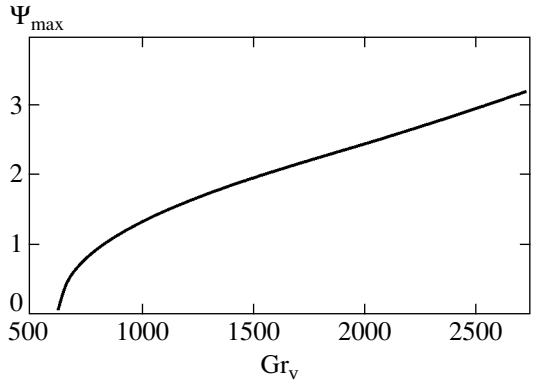


Fig. 4. Maximum value of streamfunction vs. Grashof-number modulation amplitude for ω = 21.10 (first resonance region).

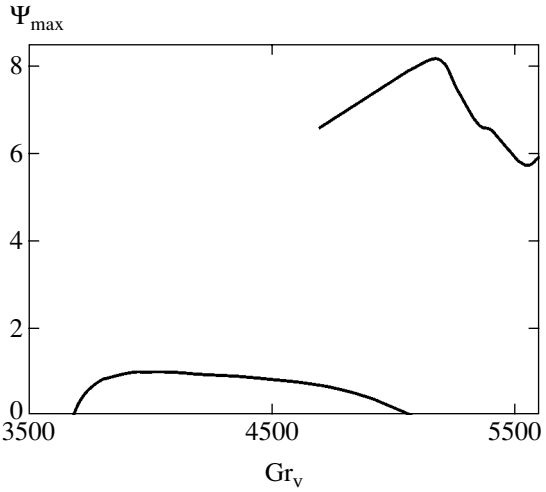


Fig. 5. Maximum value of streamfunction vs. Grashof-number modulation amplitude for ω = 7.75 (second resonance region).

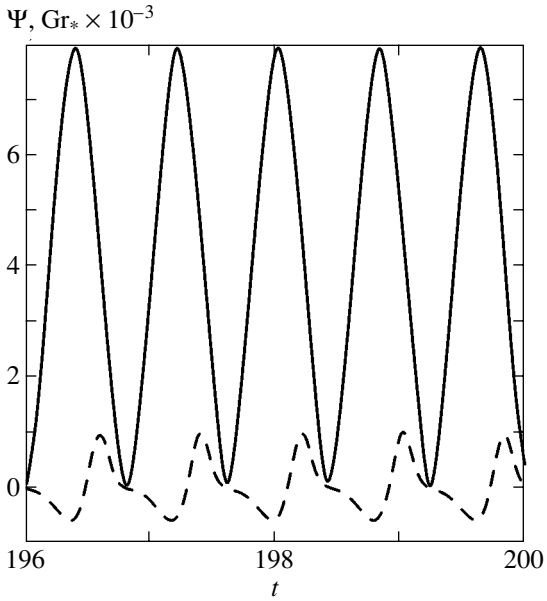


Fig. 6. Time variation of streamfunction (dashed curve) and modulated Grashof number (solid curve) at a point inside the convection cell: $Gr_v = 3950$, ω = 7.75 (second resonance region).

When $m = 1$ and ω = 21.10 (first resonance region), oscillatory perturbations with $k_m = 2.71$ evolve into a subharmonic regime of thermal vibrational convection. The variation of the streamfunction at a particular point inside a convection cell (e.g., at $x = L/4$, $z = 1/2$) is represented by the dashed curve in Fig. 2. Its period is twice the modulation period:

$$Gr_*(\omega t) = Gr + Gr_v \sin(\omega t).$$

The convection pattern is a standing wave: the streamfunction, temperature, and concentration fields oscillate about their mean values at every point. The spatial distributions of these variables at a certain instant are represented by the contour maps shown in Fig. 3. The curve of $\Psi_{max}(Gr_v)$ shown in Fig. 4 demonstrates that local bifurcation of the equilibrium state to a standing-wave pattern occurs as a result of a soft-mode instability at $Gr_v = 625$. The critical Grashof numbers predicted by linear theory and nonlinear computations differ by approximately 8%. It should be noted that the outcome of pattern formation in a modulated field is qualitatively different from that in the limit of high-frequency vibration, when a slow transient oscillatory process leads to the formation of a stationary convection pattern that exists in a subcritical region of the parameter space as well. This can be explained by the resonant effect of modulation on natural oscillatory convection in a binary mixture.

In the region of instability with respect to fundamental-frequency perturbations (corresponding to $m = 2$), the system exhibits hysteretic transitions between different convection patterns. Figure 5 shows the flow intensity (maximum value of streamfunction) versus Gr_v. Above the threshold $Gr_m = 3700$ (which differs from that predicted by linear theory by not more than 1.5%), a soft-mode instability leads to nonlinear oscillation at the gravity-modulation frequency illustrated by the dashed curve in Fig. 6, which represents the time evolution of the streamfunction at a particular point in the convection cell ($x = L/4$, $z = 1/2$). Figure 6 demonstrates that the period of nonlinear oscillation is equal to the modulation period of Gr_*. Note that the streamfunction has a nonzero mean value in this regime. With increasing modulation amplitude, the oscillation amplitude reaches a maximum and decreases to zero at $Gr_v = 5105$.

At $Gr_v = 5119$, a hard onset of oscillation at the subharmonic frequency is observed, with an amplitude several times larger than that of the fundamental-frequency mode. This regime remains stable with further increase in Gr_v.

Gradually reducing the parameter Gr_v, we observe how the oscillatory pattern of convection becomes more complicated. Figures 7a–7f show the oscillation spectra obtained by post-processing numerical data

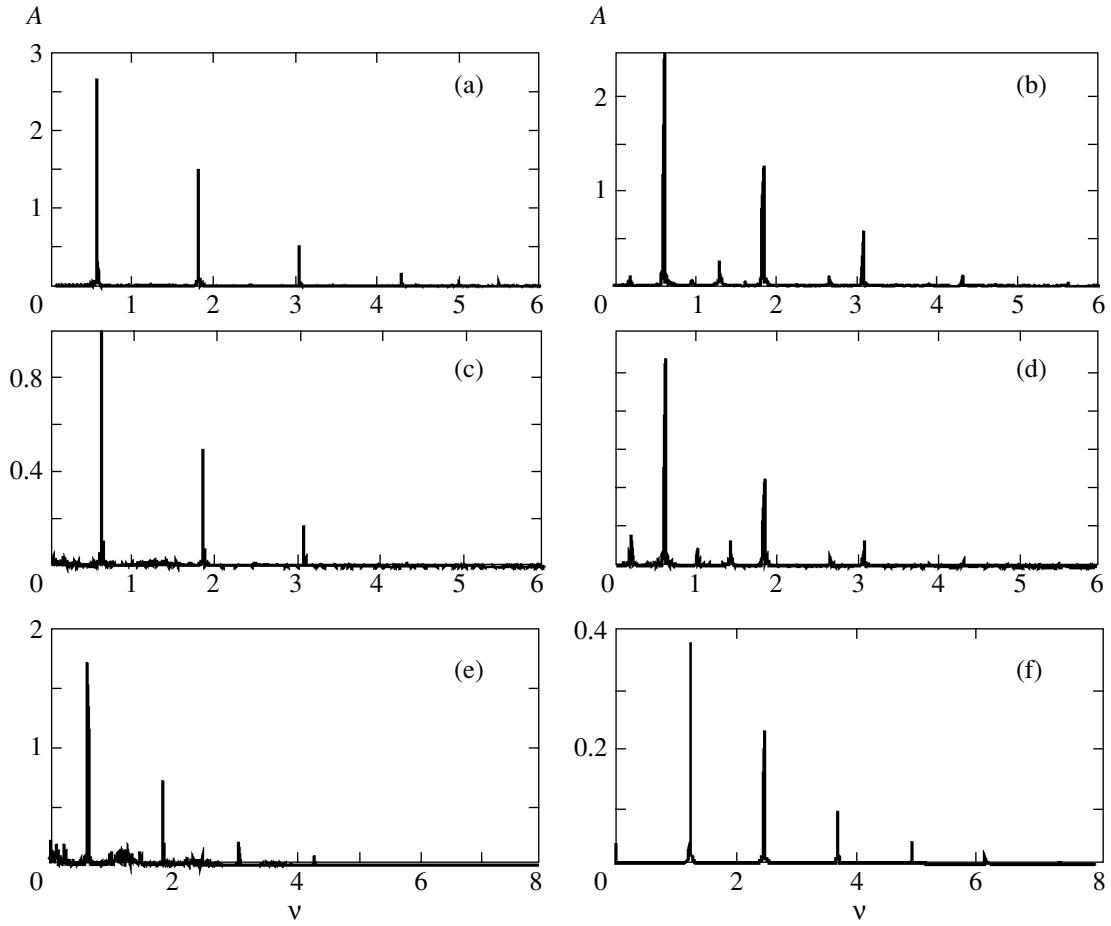


Fig. 7. Spectra of regular and chaotic oscillations in the second resonance region: Gr_v = 5612 (a), 5400 (b), 5255 (c), 5161 (d), 4692.5 (e), 4692.4 (f).

with a fast Fourier transform algorithm. The spectrum obtained for Gr_v = 5612 (Fig. 7a) contains the subharmonic mode with frequency $\nu_{1/2} = \nu/2 = 0.617$ ($\nu = \omega/2\pi$) and higher subharmonic modes with frequencies $\nu_k = k\nu_{1/2}$ (k = 1, 3, 5, …); i.e., this spectrum corresponds to a subharmonic regime of regular oscillation. At lower vibration frequencies, sidebands with frequencies $(\nu_{ki} \pm \nu_{kj})/2$ appear, as illustrated by Fig. 7b for Gr_v = 5400. In the interval 4692.5 < Gr_v < 5375, we observe chaotic oscillations with continuous spectra dominated by the half-frequency and higher subharmonic modes, as illustrated by Figs. 7c and 7e for Gr_v = 5255 and 4692.5, respectively. It should be noted here that chaotic convection is observed at relatively low supercriticality, when $Gr_v/Gr_{v0} < 1.27$. The interval of modulation amplitude indicated above contains a periodicity window where a 1/3-subharmonic oscillation mode is observed, as illustrated by the spectrum for Gr_v = 5161 shown in Fig. 7d. At Gr_v < 4692.5, the chaotic regime becomes unstable and the system oscillates at the fundamental frequency (ν = 1.234) with a smaller amplitude (see Fig. 7f). At 4692.5 < Gr_v < 5105, the resonance regions in the parameter space overlap and two convection regimes characterized by different heat transfer rates emerge and disappear via hysteretic transitions as the modulation amplitude increases or decreases.

4. CONCLUSIONS

The results obtained in this study illustrate the emergence, nonlinear evolution, and stabilization of thermally driven convection patterns in a horizontal layer of a binary mixture with negative Soret coupling subjected to vertical vibration with arbitrary frequency. It is shown that the supercritical flow regimes characteristic of the first and second resonance regions in the parameter space develop via soft-mode transitions at threshold parameter values consistent with predictions of linear stability theory. A spectral analysis of time-domain data performed to examine the behavior of nonlinear convection patterns in the second resonance region reveals diverse regular and chaotic regimes involving transitions between fundamental-frequency and subharmonic modes.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project nos. 05-01-00789 and 04-01-96029 and by CRDF, Award PE-009.

REFERENCES

1. R. W. Walden, P. Kolodner, A. Passner, and C. M. Surko, Phys. Rev. Lett. **55**, 496 (1985).
2. E. Moses and V. Steinberg, Phys. Rev. Lett. **56**, 1373 (1986).
3. P. Kolodner, Phys. Rev. A **46**, 1739 (1992).
4. E. Knobloch, Phys. Rev. A **34**, 1538 (1986).
5. O. Batiste, M. Net, I. Mercader, and E. Knobloch, Phys. Rev. Lett. **86**, 2309 (2001).
6. P. Matura, D. Jung, and M. Lücke, Phys. Rev. Lett. **92**, 254501 (2004).
7. M. I. Shliomis and B. L. Smorodin, Phys. Rev. E **71**, 036312 (2005).
8. P. L. Kapitsa, Zh. Éksp. Teor. Fiz. **21**, 588 (1951).
9. G. Ahlers, P. C. Hohenberg, and M. Lucke, Phys. Rev. A **32**, 3493 (1985).
10. G. Z. Gershuni, A. K. Kolesnikov, J.-C. Legros, and B. I. Myznikova, Int. J. Heat Mass Transf. **42**, 547 (1999).
11. G. Z. Gershuni, A. K. Kolesnikov, J.-C. Legros, and B. I. Myznikova, in *Vibration Effects in Fluid Mechanics* (Perm, 1998), p. 97 [in Russian].
12. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics*, 3rd ed. (Nauka, Moscow, 1986; Pergamon, Oxford, 1987).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 101, No. 6, 2005, pp. 1147–1152.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 6, 2005, pp. 1307–1313.

STATISTICAL, NONLINEAR, AND SOFT MATTER PHYSICS

Effect of Acceleration by Internal and External Force Fields on Particle Motion in Intermediate Regimes between the Hydrodynamic and Free-Molecular Limits

S. D. Shandakov[a,*], A. G. Nasibulin[b], Yu. I. Polygalov[a], E. Yu. Samchinskiĭ[a], and E. I. Kauppinen[b,c]

[a] *Kemerovo State University, Kemerovo, 650043 Russia*
[b] *Center for New Materials and Department of Engineering Physics and Mathematics, Helsinki University of Technology, P.O. Box 1602, FIN-02044 VTT, Espoo, Finland*
[c] *VTT Processes, Aerosol Technology Group, P.O. Box 1602, FIN-02044 VTT, Espoo, Finland*
**e-mail: cphys@kemsu.ru, Sergey.Shandakov@hut.fi*

Received June 20, 2005

Abstract—The kinetic theory of gases is applied to analyze slow translational motion of low-concentration particles driven by an external force in a homogeneous gas. The analysis takes into account the diffusion due to the difference in acceleration between particles and molecules in internal and external force fields. A general expression is derived for the particle drag force in hydrodynamic, free-molecular, and intermediate regimes. This expression reduces to a simple relation between the drag force and its values in the hydrodynamic and free-molecular limits and the force of intermolecular interaction between particles and gas molecules. In the case of spherically symmetric potential of interaction between the particle and molecules, the drag force is the harmonic mean of its limit values. © *2005 Pleiades Publishing, Inc.*

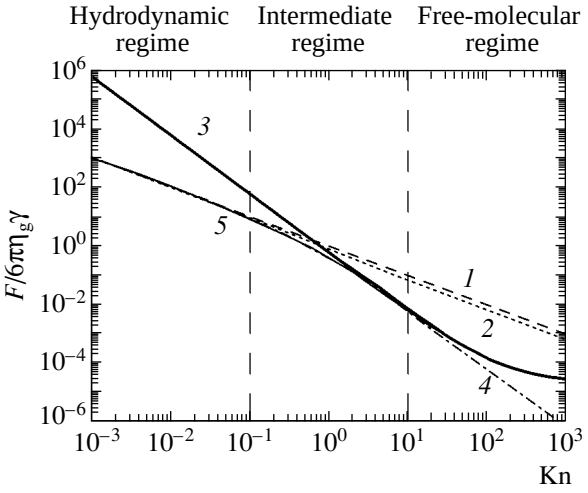
1. INTRODUCTION

The last decade has seen rapid progress in nanotechnologies (e.g., see [1–3]). The need to develop methods for synthesis and assembly of nanostructures stimulates theoretical prediction of their characteristics based on available experimental data, which obviously include characteristics of motion of nanoparticles and nanoparticle aggregates in gases [4–9]. Even though the kinetic theory of gases can be considered complete [10, 11], intermolecular interaction between polyatomic gas molecules or atoms and a particle or an ion remains an issue [10–14]. Moreover, expressions (e.g., for particle or ion drag force or mobility) obtained in the free-molecular and hydrodynamic limits of the theory are mutually inconsistent [11]. One simple method for determining the characteristics particle motion in the intermediate regime relies on the use of formal averaging procedures or correction factors involving empirical parameters. In particular, the coagulation rate in the intermediate regime has been successfully determined by using the harmonic-mean approximation [15, 16]. In the most accurate calculations of the particle drag force, a correction factor is introduced into the Stokes law [17, 18]. Despite numerous attempts to match the expressions for the drag force obtained for limit regimes, in particular, by correcting boundary conditions or introducing empirical parameters, the problem remains unsolved [11].

Generally, it is assumed that an external force field does not accelerate the motion of a particle or an ion in both free-molecular and hydrodynamic limits. In the free-molecular limit, when the gas is not perturbed by particle motion, the drag force on a particle or an ion moving in a gas with a mass-average velocity $\mathbf{V}_p$ is related to the binary diffusion coefficient by the Stokes–Einstein formula [12]

$$\mathbf{F}_{\text{diff}} = -\frac{k_B T}{D_{12}}\mathbf{V}_p, \tag{1}$$

where k_B is Boltzmann's constant and T is temperature. Note that this relation is exact only in the limit of vanishing external field [12]. The subscript "diff" in (1) refers to the free-molecular limit interpreted as the regime of diffusive particle motion (when forces are negligible). The subscript "hydr" used below refers to the hydrodynamic limit, in which the particle velocity is constant, the force acting on it is finite, and diffusion is negligible. In both limits, particle acceleration is neglected and diffusion due to the difference in acceleration between particles and molecules moving in a force field is ignored accordingly. In the intermediate regime, the acceleration of particles or ions by a force field becomes increasingly important as the particle–molecule collision frequency decreases with particle size. In this paper, we analyze particle motion in both



Drag force predicted for intermediate regimes of particle motion between hydrodynamic and free-molecular limits: (*1*) Stokes law; (*2*) Basset's formula; (*3*) Epstein's approximation; (*4*) Millikan's formula; (*5*) harmonic mean.

limit regimes and in the intermediate regime, taking into account particle acceleration by a force field.

2. AVAILABLE MODELS

In the free-molecular limit, when the perturbation of the state of the gas caused by a moving particle is negligible, the expression for the binary diffusion coefficient known from the kinetic theory of gases can be used to rewrite (1) for a sphere as follows [10]:

$$\mathbf{F}_{\text{diff}} = -\frac{8p(R+R_g)^2\Omega_{12}^{*(1,1)}}{3(k_BT/2\pi\mu)^{1/2}}\mathbf{V}_p = \frac{8\pi\eta_g(R+R_g)^2\Omega_{12}^{*(1,1)}}{3\lambda}\sqrt{\frac{\mu}{m_g}}\mathbf{V}_p, \quad (2)$$

where p is the gas pressure; $\mu = m_g m_p/(m_g + m_p)$ is the reduced mass; m_p and m_g are the particle and gas-molecule masses, respectively; R and R_g are the corresponding radii; $\Omega_{12}^{*(1,1)}$ is the reduced collision integral; and λ is the mean free path related to the absolute viscosity η_g by the formula

$$\lambda = \frac{(\pi k_BT/2m_g)^{1/2}\eta_g}{p}.$$

In the hydrodynamic limit, the motion of the gas perturbed by a moving particle is approximately described by the Stokes equations subject to the impermeability and no-slip conditions

$$U_{gn} = U_{pn}, \quad U_{g\tau} = U_{p\tau}, \quad (3)$$

where U_g and U_p are the respective gas and particle velocities, and the subscripts n and τ denote the gas-velocity components normal and tangential to the particle surface.

Available solutions to the Stokes equations for various systems have been used to obtain expressions for the drag force [19]. In particular, the drag force acting on a spherical particle executing a translational motion is given by the classical Stokes law

$$\mathbf{F}_{\text{hydr}} = -6\pi\eta_g R\mathbf{U}_p. \quad (4)$$

It can be shown that this expression (valid in the hydrodynamic limit) is inconsistent with expression (2) (valid in the free-molecular limit). In one attempt to match them, the no-slip condition was replaced with a partial slip boundary condition for the flow around a body. A solution to the Stokes equation for slip flow around a sphere was found by Basset (see [19]):

$$\mathbf{F}_{\text{hydr}} = -6\pi\eta_g R\mathbf{U}_p\frac{1+2\eta_g/R\beta}{1+3\eta_g/R\beta}, \quad (5)$$

where the slip coefficient β vanishes and goes to infinity in the case of mirror and diffuse reflection of molecules from the particle surface. A more accurate calculation (see [11]) leads to the expression

$$\mathbf{F}_{\text{hydr}} = -6\pi\eta_g R\mathbf{U}_p\frac{1+2a_+\sqrt{2/\pi}\lambda/R}{1+3a_+\sqrt{2/\pi}\lambda/R} \quad (6)$$

with $a_+ = 1.0161$.

By introducing empirical parameters A, a_1, and a_2 into the Stokes law, the following expression attributed to Cunningham, Knudsen, Weber, and Millikan has been obtained for the drag force on a spherical particle:

$$\mathbf{F} = -\frac{6\pi\eta_g R\mathbf{U}_p}{1+\text{Kn}(A+a_1\exp(-a_2/\text{Kn}))}, \quad (7)$$

where $\text{Kn} = \lambda/R$ is the Knudsen number. The values of A, a_1, and a_2 are based on experimental data obtained for small particles by Millikan with coauthors. The following values have been adopted by taking into account the results of subsequent experiments: $A = 1.257$, $a_1 = 0.40$, and $a_2 = 1.1$ [17]. Expression (7), hereinafter referred to as Millikan's formula, is generally used as a standard for validating other theoretical models [18]. However, this expression does not allow for intermolecular interaction in an intermediate regime, which strongly depends on the molecular species [11].

The figure shows results obtained by using expressions (2), (4), (6), and (7) as the dimensionless drag force versus the Knudsen number. The collision integral in (2) is represented in Epstein's approximation, $\Omega_{12}^{*(1,1)} = 1 + \alpha\pi/8$, with accommodation coefficient $\alpha = 1$. The figure demonstrates that the results calculated by using expressions (4) and (6) deviate from the

predictions based on (2) as Kn $\longrightarrow \infty$; i.e., the hydrodynamic model overestimates the force.

To match the expressions for the drag force in the hydrodynamic and free-molecular limits, we consider below possible modifications in solutions to the Stokes equation made by introducing corrections allowing for difference in molecule and particle acceleration by a force field into the gas-molecule and particle velocity distributions.

3. PARTICLE ACCELERATION BY A FORCE FIELD

We consider the effect of particle diffusion due to the difference in acceleration between particles and gas molecules by a force field on their motion in an intermediate regime. For simplicity, our analysis is restricted to slow translational motion of low-concentration particles driven by an external force in a homogeneous gas. Under this restriction, we can ignore both nonlinear terms with respect to the hydrodynamic velocity of a particle and changes in pressure, temperature, and concentrations of particles and molecules. However, this model can be applied to solve a broad variety of practical problems.

Following [11], we introduce corrections taking into account acceleration of particles and molecules by a force field into the corresponding velocity distribution functions $f_k^{(1)}$ ($k = 1$ and 2 for molecules and particles, respectively):

$$\begin{aligned} f_k^{(1)} &= f_k^{(0)}(1 + \varphi_k^{(1)}), \\ f_k^{(0)} &= n_k(\beta_k/\pi)^{3/2}\exp(-\beta_k C_k^2), \\ \mathbf{C}_k &= \mathbf{V}_k - \mathbf{U}_k, \quad \beta_k = m_k/2k_\mathrm{B}T. \end{aligned} \tag{8}$$

Here, $f_k^{(0)}$ denotes a velocity distribution functions in the zeroth approximation; $\mathbf{C}_k$ and $\mathbf{V}_k$ are the thermal and total velocities of the kth component, respectively; n_k is the concentration of the kth component; and $\varphi_k^{(1)}$ is the perturbation of the molecule or particle distribution function caused, respectively, by particle or molecule motion. For isothermal incompressible flows, this correction is written as

$$\varphi_k^{(1)} = -\frac{2\beta_k\eta_k}{p}\mathbf{C}_k\mathbf{C}_k\nabla\mathbf{U}_k - \frac{1}{n}\sum_l D_k^l(\mathbf{C}_k\cdot\mathbf{d}_1), \tag{9}$$

where n is the total concentration of mixture components, η_k is the absolute viscosity of the kth component, $D_k^l = D_k^l(C_k)$, and the thermodynamic force that drives the kth component in the absence of concentration and overall pressure gradients is expressed as

$$\mathbf{d}_k = -\frac{\rho_k}{p}\left(\frac{\mathbf{F}_k}{m_k} - \sum_l \frac{\rho_l}{\rho}\frac{\mathbf{F}_l}{m_l}\right). \tag{10}$$

For a binary mixture, when $\mathbf{d}_1 + \mathbf{d}_2 = 0$, this expression reduces to

$$\mathbf{d}_1 = -\mathbf{d}_2 = -\frac{\rho_1\rho_2}{p\rho}\left(\frac{\mathbf{F}_1}{m_1} - \frac{\mathbf{F}_2}{m_2}\right), \tag{11}$$

where m_1 and m_2 are the molecule and particle masses; ρ, ρ_1, and ρ_2 are the total, gas, and particle densities; and $\mathbf{F}_1$ and $\mathbf{F}_2$ are the forces acting on the molecule and particle, respectively. We use expression (11) to rewrite (9) as follows:

$$\begin{aligned} \varphi_k^{(1)} &= -\frac{2\beta_k\eta_k}{p}(\mathbf{C}_k)_i(\mathbf{C}_k)_j(\nabla\mathbf{U}_k)_{ij} + \frac{\rho_1\rho_2}{pn\rho} \\ &\times (D_k^1 - D_k^2)\left(\frac{1}{m_1}(\mathbf{C}_k\cdot\mathbf{F}_1) - \frac{1}{m_2}(\mathbf{C}_k\cdot\mathbf{F}_2)\right), \end{aligned} \tag{12}$$

where summation over repeated indices is assumed; i, $j = 1$, 2, and 3 denote vector components in a coordinate system ($\mathbf{n}$, $\boldsymbol{\tau}_1$, $\boldsymbol{\tau}_2$).

The function $D_k^l = D_k^l(C_k)$ is defined in terms of diffusion coefficients $\mathbf{D}_{kl}$ by an integral relation [11]:

$$\mathbf{D}_{kl} = \frac{1}{3nn_k}\int f_k^{(0)}C_k^2 D_k^l(C_k)d\mathbf{V}_k. \tag{13}$$

According to [11], the coefficients D_{kl} for a binary mixture are related to D_{12} can be represented as

$$\begin{aligned} \mathbf{D}_{11} &= \frac{\rho_2}{\rho_1}\frac{m_1m_2}{(\rho/n)^2}D_{12}, \quad \mathbf{D}_{22} = \frac{\rho_1}{\rho_2}\frac{m_1m_2}{(\rho/n)^2}D_{12}, \\ \mathbf{D}_{12} &= \mathbf{D}_{21} = -\frac{m_1m_2}{(\rho/n)^2}D_{12}. \end{aligned} \tag{14}$$

Equation (12) is derived without taking into account internal degrees of freedom of molecules and particles. Even though difficult problems arise in analyzing the effects due to internal degrees of freedom and configuration of interacting particles, polyatomic gases can be described by methods developed in the kinetic theory of gases [10, 11]. In these methods, description of the dependence of a velocity distribution function on molecular or particle rotation characteristics is reduced to representation of transport coefficients as functions of these characteristics. Following this approach, we apply expression (12) to polyatomic molecules and particles, assuming that the influence of rotation character-

istics can be described by formulas for viscosity and diffusion coefficients.

The expression for the drag force in the hydrodynamic limit is obtained by assuming that the particle velocity is determined by its hydrodynamic velocity. Accordingly, we calculate the particle velocity averaged over molecular and particle velocities and volume, using the above particles and gas-molecule distribution functions modified by taking into account the acceleration of particles and gas molecules by a force field:

$$\langle V_\mathrm{p}\rangle = \frac{1}{n_\mathrm{p}n_\mathrm{g}} \times\left\langle \int\int_{\mathbf{V}_\mathrm{g}\mathbf{V}_\mathrm{p}} (\mathbf{n}_z\cdot\mathbf{V}_\mathrm{p}) f_\mathrm{p}^{(0)} f_\mathrm{g}^{(0)}(1+\varphi_\mathrm{g}^{(1)}+\varphi_\mathrm{p}^{(1)})d\mathbf{V}_\mathrm{p}d\mathbf{V}_\mathrm{g}\right\rangle_v. \quad (15)$$

Here, the angle brackets denote averaging over a volume v. Since the direction of translational motion of particles is parallel to the external force, expression (15) is written in scalar form with $\mathbf{n}_z$ denoting the unit vector in the direction of the external force. We rewrite this expression as

$$\langle V_\mathrm{p}\rangle = U_\mathrm{p} + U_\mathrm{diff}, \quad (16)$$

where U_p is the hydrodynamic velocity. The diffusion velocity U_diff depends on the difference in acceleration between molecules and particles by a force field:

$$U_\mathrm{diff} = U_\mathrm{diff}^\mathrm{ext} + U_\mathrm{diff}^\mathrm{int}, \quad (17)$$

$$U_\mathrm{diff}^\mathrm{ext} = -\sum_{k=1}^{2}\frac{\rho_1\rho_2}{pm_2n\rho} \times\left\langle \int\int_{\mathbf{V}_\mathrm{g}\mathbf{V}_\mathrm{p}} (\mathbf{n}_z\cdot\mathbf{V}_\mathrm{p})\frac{f_\mathrm{p}^{(0)}f_\mathrm{g}^{(0)}}{n_\mathrm{p}n_\mathrm{g}}(D_k^2-D_k^1)(\mathbf{C}_k\cdot\mathbf{F})d\mathbf{V}_\mathrm{p}d\mathbf{V}_\mathrm{g}\right\rangle_v, \quad (18)$$

$$U_\mathrm{diff}^\mathrm{int} = -\sum_{k=1}^{2}\frac{\rho_1\rho_2}{p\mu n\rho} \times\left\langle \int\int_{\mathbf{V}_\mathrm{g}\mathbf{V}_\mathrm{p}} (\mathbf{n}_z\cdot\mathbf{V}_\mathrm{p})\frac{f_\mathrm{p}^{(0)}f_\mathrm{g}^{(0)}}{n_\mathrm{p}n_\mathrm{g}}(D_k^2-D_k^1)(\mathbf{C}_k\cdot\mathbf{F}_\mathrm{int})d\mathbf{V}_\mathrm{p}d\mathbf{V}_\mathrm{g}\right\rangle_v, \quad (19)$$

$$\mathbf{F}_\mathrm{int} = -\frac{m_2}{\mu}\int_v n_\mathrm{g}\mathbf{F}_1 dv = -\frac{m_2}{\mu}\langle\mathbf{F}_1\rangle_v, \quad (20)$$

where the diffusion velocities $U_\mathrm{diff}^\mathrm{ext}$ and $U_\mathrm{diff}^\mathrm{int}$ are associated with the external and internal forces acting on particles. Expression (18) is written by assuming that the external force acts only on particles; for particle moving with constant velocities, it is counterbalanced by the drag force ($\mathbf{F}_2 = -\mathbf{F}_1$). Expressions (19) and (20) are written by using the fact that the internal forces $\langle\mathbf{F}_2\rangle_v$ exerted on the particle by surrounding molecules are counterbalanced the respective forces $\langle\mathbf{F}_1\rangle_v$ exerted by the particle on the molecules:

$$\langle\mathbf{F}_1\rangle_v = -\langle\mathbf{F}_2\rangle_v.$$

The force $\mathbf{F}_1$ contained in expression (20) is exerted by the particle on a gas molecule. Since the particle concentration is assumed to be low, interparticle interactions are not taken into account.

Using (13) to calculate the integral of expression (18) with respect to velocities, we obtain

$$U_\mathrm{diff}^\mathrm{ext} = -\frac{\rho_1\rho_2}{p\rho}(\mathbf{D}_{22}-\mathbf{D}_{21})F. \quad (21)$$

Substituting (14) and using the expression $p = nk_\mathrm{B}T$, we rewrite it as

$$U_\mathrm{diff}^\mathrm{ext} = -(\rho_1-\rho_2)\frac{nm_1\rho_1}{\rho^3}\frac{D_{12}}{k_\mathrm{B}T}F. \quad (22)$$

When the particle concentration in the gas is low, this expression reduces to

$$U_\mathrm{diff}^\mathrm{ext} = -\frac{D_{12}}{k_\mathrm{B}T}F. \quad (23)$$

To calculate diffusion velocity (19), we must know certain characteristics of the forces acting on molecules and particles. In the case of central forces, the volume-averaged accelerations of molecules and particles vanish; i.e., these forces do not contribute to the diffusion velocity. Otherwise, we have to take into account the asymmetry of internal forces.

Using (13) and (14) to calculate the integral of expression (19) with respect to velocities, we obtain

$$U_\mathrm{diff}^\mathrm{int} = (\rho_1-\rho_2)\frac{nm_1\rho_1}{\rho^3}\frac{D_{12}}{k_\mathrm{B}T}F_{\mathrm{int},z}, \quad (24)$$

where $F_{\mathrm{int},z}$ is the component of the force exerted on the particle by surrounding molecules parallel to the external force. For low-concentration particles, expression (24) reduces to

$$U_\mathrm{diff}^\mathrm{int} = \frac{D_{12}}{k_\mathrm{B}T}F_{\mathrm{int},z}. \quad (25)$$

According to expressions (24) and (25), if the force $\mathbf{F}_\mathrm{int}$ exerted by gas molecules on the particle is asymmetric, then its volume-averaged value is finite. For example, if the center of charge of a particle moving in external electric field falls behind its center of mass, then the force exerted on the particles by the molecules

polarized by the field generated by the particle is parallel to the direction of its motion, and so is its diffusion velocity. When both interaction potential and particle's orientation are known, the mean value of $\mathbf{F}_{\text{int}}$ can be represented in explicit form. In particular, if the charge distribution is known for a charged particle, then we can find the ion–dipole interaction force between the particle and neutral molecules polarized by the field generated by the particle [13].

Using (23) and (25), we rewrite expression (16) as

$$\langle V_{\text{p}} \rangle = U_{\text{p}} - \frac{D_{12}}{k_{\text{B}}T}F + \frac{D_{12}}{k_{\text{B}}T}F_{\text{int},z}. \tag{26}$$

According to (16) and (26), the mean particle velocity is a linear combination of its hydrodynamic and diffusion velocities. Therefore, the mean velocity $\langle V_{\text{p}} \rangle$ (interpreted as hydrodynamic velocity) in formulas derived in the hydrodynamic limit should be replaced by the velocity U_{p} determined from (26):

$$F = \frac{F_{\text{hydr}}}{\langle V_{\text{p}} \rangle}\left(\langle V_{\text{p}} \rangle - \frac{D_{12}}{k_{\text{B}}T}F + \frac{D_{12}}{k_{\text{B}}T}F_{\text{int},z} \right). \tag{27}$$

Using Stokes–Einstein formula (1), we rewrite this expression as

$$F = F_{\text{hydr}}\frac{F_{\text{diff}} + F_{\text{int},z}}{F_{\text{diff}} + F_{\text{hydr}}}. \tag{28}$$

Thus, when the acceleration of particles and molecules by a force field is taken into account, the drag force can be expressed in terms of its values in the hydrodynamic and free-molecular limits and the force of interaction between a particle and surrounding gas molecules. The value of the force in the hydrodynamic limit used in this expression is determined without allowance for the effects of diffusion and slip. The latter effect is taken into account in the expression for the force in the free-molecular limit in terms of the binary diffusion coefficient.

In the case of a central interaction potential, the drag force can be represented as the harmonic mean of the drag forces determined in the free-molecular and hydrodynamic limits:

$$\frac{1}{F} = \frac{1}{F_{\text{diff}}} + \frac{1}{F_{\text{hydr}}}. \tag{29}$$

The results of calculations using expression (29) shown in the figure demonstrate that the values of the drag force given by the harmonic mean and Millikan's formula are virtually equal in both hydrodynamic and intermediate regimes. The discrepancy between these predictions at high Knudsen numbers can be explained by the neglected interaction between particles and molecules.

Expressions (28) and (29) are obtained by representing the drag force in terms of the mean particle velocity, which is a linear combination of the limit velocities. Therefore, expressions analogous to those presented above may be valid for other characteristics determined by particle velocities. In particular, particle coagulation rate can be calculated as a harmonic mean [15, 16] by assuming its proportionality to the frequency of collisions between particles, which is determined by the their relative velocity.

Thus, it is shown here that the harmonic mean is a theoretically well-grounded approximation applicable when particle–molecule interactions can be described by a spherically symmetric potential.

Expressions (28) and (29) determine the particle drag force when its values in the hydrodynamic and free-molecular limits are known. Since particle dynamics in the limit regimes have been well studied for particles of various shapes, these expressions can be used to calculate the drag force for particles of any size, including nanoparticles.

Note that the free-molecular regime is interpreted here as particle motion driven by an external force that does not perturb the state of the gas; i.e., the force is set to zero in the free-molecular limit. This approach is used only to determine the diffusion coefficient defined without taking into account the diffusive fluxes due to difference in acceleration between molecules and particles. The validity of the diffusion coefficients obtained in the free-molecular limit for large particles is questionable because of limited applicability of Boltzmann's equation to kinetics of large particles. However, methods other than the classical one (Chapman–Enskog method) can be applied to solve Boltzmann's equation. In particular, the probability of collisions between molecules of the same species can be higher than the collision probability between molecules of different species [20]. A solution based on this model can be applied to particles of size much greater than the mean free path. When the gas is treated as isothermal and incompressible, this solution reduces to the Chapman–Enskog approximation. Thus, we can apply the formulas obtained for the drag force F_{diff} in the free-molecular limit, at least, to nanoparticles.

4. CONCLUSIONS

The kinetic theory of gases is applied to analyze the effect of external and internal forces on slow translational motion of low-concentration particles in a homogeneous gas. The expressions for the drag force obtained in the hydrodynamic limit are matched with those obtained in the free-molecular limit by introducing a correction allowing for difference in molecule and particle acceleration by a force field. The simple expression presented in this paper is a simple relation between the drag force and its values in the hydrodynamic and free-molecular limits and the force of inter-

molecular interaction between particles and gas molecules. When this interaction can be described by a spherically symmetric potential, the drag force is the harmonic mean of its limit values. The results of this study make it possible to calculate the drag force for particles of any size by using known expressions obtained for the limit regimes.

ACKNOWLEDGMENTS

This work was supported, in part, by the Academy of Sciences of Finland and European Union, grant no. HPRN-CT-2002-00328. One of us (S.D.S.) gratefully acknowledges the support provided by the European Commission (Marie Curie Incoming International Fellowship 022110-2005).

REFERENCES

1. L. Fu, L. Cao, Y. Liu, and D. Zhu, Adv. Colloid Interface Sci. **111**, 133 (2004).
2. M. Pitkethly, Materials Today **6**, 36 (2003).
3. Y. H. D. Pui and D. Chen, J. Aerosol Sci. **28**, 539 (1997).
4. J. Fernández de la Mora, L. de Juan, T. Eichler, and J. Rosell, TrAC Trends Anal. Chem. **17**, 328 (1998).
5. S. Ude and J. Fernández de la Mora, J. Aerosol Sci. **34**, 1245 (2003).
6. A. G. Nasibulin, E. I. Kauppinen, B. A. Thomson, and J. Fernández de la Mora, J. Nanopart. Res. **4**, 449 (2002).
7. V. Ya. Rudyak and S. L. Krasnolutskiĭ, Dokl. Akad. Nauk **381**, 623 (2001) [Dokl. Phys. **46**, 897 (2001)].
8. V. Ya. Rudyak, S. L. Krasnolutskiĭ, A. G. Nasibulin, and E. I. Kauppinen, Dokl. Akad. Nauk **386**, 624 (2002) [Dokl. Phys. **47**, 758 (2002)].
9. A. Moisala, A. G. Nasibulin, S. D. Shandakov, *et al.*, Carbon **43**, 2066 (2005).
10. S. Chapman and T. G. Cowling, *Mathematical Theory of Non-Uniform Gases*, 3rd ed. (Cambridge Univ. Press, Cambridge, 1990; Inostrannaya Literatura, Moscow, 1960).
11. J. H. Ferziger and H. G. Kaper, *Mathematical Theory of Transport Processes in Gases* (North-Holland, Amsterdam, 1972; Mir, Moscow, 1976).
12. E. W. McDaniel and E. A. Mason, *The Mobility and Diffusion of Ions in Gases* (Wiley, New York, 1973; Mir, Moscow, 1976).
13. S. D. Shandakov, A. G. Nasibulin, and E. I. Kauppinen, J. Aerosol Sci. **36**, 1125 (2005).
14. Z. Li and H. Wang, Phys. Rev. E **68**, 061206 (2003).
15. S. E. Pratsinis, J. Colloid Interface Sci. **124**, 416 (1988).
16. F. O. Ernst and S. E. Pratsinis, submitted to J. Aerosol Sci.
17. N. A. Fuks, *Mechanics of Aerosols* (Akad. Nauk SSSR, Moscow, 1955) [in Russian].
18. B. K. Annis, A. P. Malinauskas, and E. A. Mason, J. Aerosol Sci. **3**, 55 (1972).
19. J. Happel and H. Brenner, *Low Reynolds Number Hydrodynamics* (Prentice-Hall, Englewood Cliffs, N.J., 1965; Mir, Moscow, 1976).
20. V. V. Struminskiĭ, Prikl. Mat. Mekh. **38**, 203 (1974).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 101, No. 6, 2005, pp. 1153–1159.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 6, 2005, pp. 1314–1320.

STATISTICAL, NONLINEAR, AND SOFT MATTER PHYSICS

Relativistic Generalization of Quasi-Chaplygin Equations

V. P. Vlasov

Nuclear Fusion Institute, Russian Research Centre Kurchatov Institute, pl. Kurchatova 1, Moscow, 123182 Russia
e-mail: vlasov@nfi.kiae.ru

Received August 3, 2005

Abstract—A relativistic generalization of quasi-Chaplygin (quasi-gas) equations describing the evolution of unstable media with negative compressibility is proposed. Examples of the media whose dynamics can be described by the proposed equations are considered. An analytic solution to these nonlinear equations is obtained for the 1D case. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Nonlinear evolution of many unstable media can be described in the long-wave approximation by quasi-Chaplygin equations (QCEs)

$$\frac{d\mathbf{v}}{dt} = c_0^2 m \nabla \rho_*^{1/m}, \quad \frac{d}{dt} = \frac{\partial}{\partial t} + (\mathbf{v}\nabla), \tag{1.1}$$

$$\frac{d\rho_*}{dt} + \rho_*(\nabla \mathbf{v}) = 0, \tag{1.2}$$

where ρ_* is the effective density (dimensionless quantity), $\mathbf{v}$ is the velocity, c_0 is the velocity of "sound" and m is the parameter referred to as the azimuth number. These equations differ from the ideal gas equations only in that they contain negative compressibility. However, this difference is essential since, instead of running waves inherent in conventional gases, standing perturbations growing in time become prevalent. Such instabilities are encountered in nature quite often; for instance, about 50 such examples are considered in [1, 2], where a general theory of the QCEs is formulated. The properties of the system of QCEs were studied systematically later in [3]. Media whose evolution is described by QCEs are usually referred to as quasi-Chaplygin or quasi-gaseous media. The existence of analytic solutions to nonlinear equations (1.1), (1.2) is a considerable advantage of these equations both in the 1D case and 2D (stationary) case. Consequently, generalizations of these equations for which analytic solutions exist are of interest for describing specific physical processes as well as for testing numerical methods employed for solving similar nonlinear equations belonging to the class of incorrect problems for which small-scale perturbations that grow the most rapidly play a decisive role.

In this paper, we propose a relativistic generalization of QCEs (1.1) and (1.2) and demonstrate its relationship with a number of physical problems. In the 1D case, these equations have the form

$$\left(\gamma\frac{\partial}{\partial\tau} + u\frac{\partial}{\partial z}\right)y = m\left(u\frac{\partial}{\partial\tau} + \gamma\frac{\partial}{\partial z}\right)\frac{c_0^2}{c^2}\rho_*^{1/m}, \tag{1.3}$$

$$\left(\gamma\frac{\partial}{\partial\tau} + u\frac{\partial}{\partial z}\right)\ln\rho_* = -\left(u\frac{\partial}{\partial\tau} + \gamma\frac{\partial}{\partial z}\right)y. \tag{1.4}$$

The time coordinate in these equations is $\tau = ct$, c is the velocity of light,

$$\gamma = (1 - v^2/c^2)^{-1/2} = \cosh y$$

is the Lorentz parameter, z is the coordinate, and $u = \gamma v/c = \sinh y$ is the spatial component of the four-dimensional velocity; the meaning of rest of the notation remains unchanged. Equations (1.3) and (1.4) will be referred to as relativistic quasi-Chaplygin equations (RQCE) since these equations are transformed into one-dimensional equations (1.1) and (1.2), respectively, in the nonrelativistic limit ($v/c \longrightarrow 0$). In this generalization, we proceeded from Eqs. (1.3) and (1.4) derived earlier in [4] for $m = -1$, which describe the dynamics of a plasma in a relativistic skinned current pinch in the "narrow channel" approximation in zero longitudinal magnetic field. The plasma was treated as nonrelativistic in its own coordinate system and was described using the conventional isentropic equation with an exponent of 5/3. These equations (for $m = -1$) were used in [4, 5] for analytically calculating the energy spectrum of particles accelerated in a pinch when sausage-type instabilities grow in it. It was found that this spectrum successfully describes the energy spectrum of galactic ultrarays in the entire range of observed energies, which led to the hypothesis on generation of galactic ultrarays in cosmic current pinches.

Equations (1.3) and (1.4) can be derived from the equations of motion and continuity in relativistic fluid dynamics [6]:

$$wu^k\frac{\partial u_i}{\partial x^k} = \frac{\partial p}{\partial x^i} - u_i u^k\frac{\partial p}{\partial x^k}, \quad (1.5)$$

$$\frac{\partial (nu^i)}{\partial x^i} = 0, \quad (1.6)$$

where $w = e + p$ is the enthalpy; $e = \rho c^2$, ρ, p, and n are the mass density, pressure, and number density n in the intrinsic frame of reference; and x^k, u^k, and u_k are the 4-vectors of coordinates and velocity, $u^k = (\gamma, \gamma\mathbf{v}/c)$. Three spatial components of Eq. (1.5) form a relativistic generalization of the Euler equation, while the temporal component is the consequence of the former three components (the scalar product of the velocity vector and the vector equation of motion ($i = 1, 2, 3$) leads to Eq. (1.5) for $i = 0$). Equation (1.6) is a continuity equation.

To derive Eqs. (1.3) and (1.4) from Eqs. (1.5) and (1.6), we assume that

$$e = \rho_* c^2, \quad p = -\frac{c_0^2 m}{1+m}\rho_*^{1+1/m} \quad (1.7)$$

for quasi-Chaplygin media with $m \neq -1$ and $p = -c_0^2 \ln\rho_*$ for $m = -1$. The above expressions for "pressure" p can be derived from Eq. (1.1) by multiplying it by ρ_* and writing its right-hand side in the conventional form $-\nabla p$. It should be noted that quantities e and p in formulas (1.7) have dimensions of the square of velocity. Henceforth, we will disregard pressure in the formula for enthalpy, setting $w \approx \rho_* c^2$. Substituting Eq. (1.7) into (1.5), we obtain the relativistic quasi-Chaplygin equation of motion,

$$\gamma\frac{d}{dt}(\gamma\mathbf{v}) = -\nabla p_{\text{ef}} - \left(\frac{\gamma}{c}\right)^2\mathbf{v}\frac{d}{dt}p_{\text{ef}}, \quad (1.8)$$

where $p_{\text{ef}} = -c_0^2 m\rho_*^{1/m}$. In the 1D case, this equation has the form of Eq. (1.3).

Let us now consider the continuity equation (1.6). Substituting ρ_* for n in this equation leads to Eq. (1.4). It should be noted that this substitution automatically implies that "effective" density ρ_* in Eq. (1.6) pertains to the intrinsic frame of reference; i.e., under the Lorentz transformation, this quantity behaves as conventional density.

Concluding the section, we can make a few brief remarks concerning the generalization of quasi-Chaplygin equations to the case of strong gravitational fields, i.e., the form of the QCE in the general theory of relativity. Since the RQCE were derived from Eqs. (1.5) and (1.6) of relativistic fluid dynamics, the above-mentioned generalization of the QCE should be carried out in accordance with the algorithm employed in relativistic fluid dynamics. For this purpose, conventional derivatives in Eqs. (1.5) and (1.6) should be replaced by covariant derivatives (see [6]) and then relations (1.7) should be used.

2. REDUCTION OF NONLINEAR EQUATIONS (1.3) AND (1.4) TO LINEAR EQUATIONS

To solve the "1D" RQCE, we will employ the locus technique, according to which a transition should be made to reciprocal functions $z = z(x, y)$ and $\tau = \tau(x, y)$ with the variable $x = (c_0/c)^2\rho_*^{1/m}$. Differentiating these functions with respect to τ and z, we obtain the expressions for the derivatives,

$$\begin{aligned}\frac{\partial y}{\partial \tau} &= \frac{1}{J}\frac{\partial z}{\partial x}, \quad \frac{\partial x}{\partial \tau} = -\frac{1}{J}\frac{\partial z}{\partial y},\\ \frac{\partial y}{\partial z} &= -\frac{1}{J}\frac{\partial \tau}{\partial x}, \quad \frac{\partial x}{\partial z} = \frac{1}{J}\frac{\partial \tau}{\partial y},\end{aligned} \quad (2.1)$$

where

$$J = \frac{\partial \tau}{\partial y}\frac{\partial z}{\partial x} - \frac{\partial \tau}{\partial x}\frac{\partial z}{\partial y}$$

is the Jacobian of the transition. Substituting Eqs. (2.1) into RQCEs (1.3) and (1.4), we arrive at the equations

$$\begin{aligned}\gamma\frac{\partial z}{\partial x} - u\frac{\partial \tau}{\partial x} &= m\left(-u\frac{\partial z}{\partial y} + \gamma\frac{\partial \tau}{\partial y}\right),\\ \gamma\frac{\partial z}{\partial y} - u\frac{\partial \tau}{\partial y} &= \frac{x}{m}\left(u\frac{\partial z}{\partial x} - \gamma\frac{\partial \tau}{\partial x}\right).\end{aligned} \quad (2.2)$$

Then we introduce coordinate z' and time $\tau' = ct'$ of an event in the intrinsic frame of reference (in which the volume element under investigation is at rest). These quantities are related to coordinate z and time τ of the same even in the laboratory reference frame via the Lorentz transformation:

$$z' = \gamma z - u\tau, \quad \tau' = \gamma\tau - uz. \quad (2.3)$$

Differentiating functions $z'(x, y)$ and $\tau'(x, y)$ with respect to variables x and y and taking into account relations (2.2), we obtain the expressions

$$\frac{\partial z'}{\partial y} = -\left(\tau' + \frac{x}{m}\frac{\partial \tau'}{\partial x}\right), \quad \frac{\partial z'}{\partial x} = m\left(z' + \frac{\partial \tau'}{\partial y}\right), \quad (2.4)$$

which lead to the following equation:

$$x\frac{\partial^2 \tau'}{\partial x^2} + (1 + m - xm)\frac{\partial \tau'}{\partial x} + m^2\left(\frac{\partial^2 \tau'}{\partial y^2} - \tau'\right) = 0. \quad (2.5)$$

Let us first consider this equation in the nonrelativistic approximation ($v/c \longrightarrow 0$), in which $y \approx \varepsilon\eta$, where $\eta = v/c_0$ and $\varepsilon = c_0/c \ll 1$. Passing in Eq. (2.5) to variables η and $r = x^{1/2}/\varepsilon = \rho_*^{1/2m}$ and proceeding to the limit $\varepsilon \longrightarrow 0$, we arrive at the Darboux equation for function $\tau(x, y)$, which is transformed into τ' in the nonrelativistic approximation:

$$\frac{\partial^2\tau}{\partial r^2} + \frac{(1+2m)}{r}\frac{\partial\tau}{\partial r} + 4m^2\frac{\partial^2\tau}{\partial\eta^2} = 0. \tag{2.6}$$

For $m = -1/2$, Eq. (2.6) is a 2D Laplace equation, while for other values of m it can be reduced, in accordance with [1], to a 3D Laplace equation. From the multitude of solutions to the Laplace equation, we will be interested only in solutions describing perturbations in unstable media, which vanish in the limit $t \longrightarrow -\infty$. In the general theory of the QCEs, such perturbations and solutions are referred to as spontaneous. Their evolution occurs smoothly, without breaking or sharpening of profiles typical of nonlinear systems without dissipations.

Spontaneous solutions were obtained in [1, 2] using a method based on analogy with electrostatic problems. In this analogy, function $\tau(r, \eta)$ is treated as an electrostatic potential and the unperturbed state of the system ($\rho_* = 1$, $v = 0$) corresponds to point $r_0 = 1$, $\eta_0 = 0$ in the r, η space. The "electrostatic" potential $\tau(r, \eta)$ possesses the required singularity $\tau \longrightarrow -\infty$ at this point if we place "electric charges" at this point. Analytically, this is reduced to substitution of the density of these charges for zero on the right-hand side of Eq. (2.6), i.e., to a transition from the Laplace equation to the Poisson equation.

The same substitution should also be performed in the general equation (2.5) by writing it in the form

$$x\frac{\partial^2\tau'}{\partial x^2} + (1+m-xm)\frac{\partial\tau'}{\partial x} + m^2\left(\frac{\partial^2\tau'}{\partial y^2} - \tau'\right) = \sigma(x, y), \tag{2.7}$$

where $\sigma(x, y)$ is the density of the "charges" concentrated at point $x_0 = \varepsilon^2$, $y_0 = 0$, which corresponds to the unperturbed state of the medium ($\rho_* = 1$, $v = 0$). In this case, solutions to Eq. (2.7) are spontaneous solutions of the RQCEs.

3. THE GREEN FUNCTION OF RELATIVISTIC QUASI-CHAPLYGIN EQUATIONS

Before writing the Green function of Eq. (2.7), we will write particular solutions to homogeneous equation (2.5). For positive azimuthal numbers $m > 0$, these solutions have the form

$$\tau'(x, y) = L_n^m(xm)\exp(-|y|q_n), \quad q_n = \sqrt{1 + \frac{n}{m}}, \tag{3.1}$$

where $L_n^m(xm)$ are Laguerre polynomials. For negative azimuthal numbers $m < 0$, particular solutions have a different form:

$$\tau'(x, y) = h(x)L_n^{|m|}(x|m|)\exp(-|y|q_n), \quad h(x) = x^{|m|}\exp(-x|m|), \quad q_n = \sqrt{1 + \frac{1+n}{|m|}}. \tag{3.2}$$

We will seek the Green function G of Eq. (2.7) ($\sigma = \delta(x - x_0)\delta(y - y_0)$) in the form of a series in the Laguerre polynomials, which form a complete set of functions. As a result of simple transformations, we obtain an expression for G, which is applicable both for $m > 0$ and for $m < 0$; this equation has the form

$$G(\zeta, y, \zeta_0, y_0) = B(\zeta, \zeta_0) \times \sum_{n=0}^{\infty}\frac{n!L_n^{|m|}(\zeta)L_n^{|m|}(\zeta_0)}{q_n\Gamma(n + |m| + 1)}\exp(-|y - y_0|q_n), \tag{3.3}$$

where $\zeta = x|m| = \varepsilon^2 r^2|m|$, Γ is the gamma function, and quantities q_n are defined by formulas (3.1) and (3.2). For $m > 0$, function $B(\zeta, \zeta_0)$ depends only on ζ_0: $B(\zeta, \zeta_0) = b(\zeta_0)$, where

$$b(\zeta_0) = -(2|m|)^{-1}\zeta_0^{|m|}\exp(-\zeta_0),$$

and function $B(\zeta, \zeta_0)$ for $m < 0$ is a function of ζ alone: $B(\zeta, \zeta_0) = b(\zeta)$.

Let us pass to the nonrelativistic case ($v/c \longrightarrow 0$). For $\varepsilon = c_0/c \longrightarrow 0$, we replace sum (3.3) by an integral. Taking into account the formula

$$L_n^{|m|}(\zeta) \approx \left(\frac{n}{\zeta}\right)^{|m|/2} J_{|m|}(2\sqrt{n\zeta}),$$

which is valid for $\zeta \longrightarrow 0$ and $n \gg |m|$ (see [7]), we can write this integral in the form

$$G(r, r_0, u', u_0') = -\frac{1}{2|m|\varepsilon}\left(\frac{r_0}{r}\right)^m \times \int_0^{\infty} J_{|m|}(qr)J_{|m|}(qr_0)\exp(-|u' - u_0'|q)dq, \tag{3.4}$$

where

$$r = \rho_*^{1/2m}, \quad u' = \frac{v}{2|m|c_0}.$$

This formula is valid both for $m > 0$ and for $m < 0$; it should be noted that the exponent of factor $(r_0/r)^m$ contains azimuthal number m and not its absolute value. The integral in formula (3.4) can be expressed in terms of the second-order Legendre function $Q_{|m|-1/2}$ [8]:

$$\int_0^\infty J_{|m|}(qr)J_{|m|}(qr_0)\exp(-|u'-u_0'|q)dq = \frac{1}{\pi\sqrt{r_0 r}}Q_{|m|-1/2}\left(\frac{(u'-u_0')^2+r^2+r_0^2}{2r_0 r}\right). \quad (3.5)$$

Let us multiply the Green function by "charge" $e_* = 2\pi c_0 t_* |m|$, where the positive constant t_* has the dimension of time (it will be shown in the next section that $2\pi c_0 t_*$ is the wavelength of a perturbation along the z axis). Using Eqs. (3.4) and (3.5) in the nonrelativistic case and the values $u_0' = 0$ and $r_0 = 1$ for the unperturbed state of the medium, we obtain the formula given in [1, 2],

$$\frac{t}{t_*} = -r^\nu Q_{|m|-1/2}(\chi), \quad (3.6)$$

where

$$\nu = -m-\frac{1}{2}, \quad \chi = \frac{u'^2+r^2+1}{2r}.$$

Expression (3.4) can be derived directly from Eq. (2.6). Its particular solutions have the form

$$\tau = r^{-m}J_m(qr)\exp(\pm qu'),$$

where $J_m(qr)$ is the Bessel function and $u' = \eta/2|m| = v/2|m|c_0$. Consequently, to derive the Green function (3.4) from Eq. (2.6), we must carry out the substitution $0 \longrightarrow \delta(r-r_0)\delta(\eta-\eta_0)$ on the right-hand side of Eq. (2.6) and then use the familiar Fourier–Bessel integral transformation (Hankel transformation):

$$f(r,u') = \int_0^\infty f(q,u')J_{|m|}(qr)dq, \qquad \delta(r-r_0) = r_0\int_0^\infty qJ_{|m|}(qr)J_{|m|}(qr_0)dq. \quad (3.7)$$

This leads to the previous result (3.4)

4. PARAMETRIC COORDINATE REPRESENTATION

The formula for coordinate z' can be obtained by integrating Eqs. (2.4) into which we must substitute the expression $\tau' = e_* G$. Let us first perform these calculations for the linear stage of evolution of perturbation, for which $u' \longrightarrow 0$ and $r \longrightarrow 1$ so that formula (3.6) can be reduced to

$$\frac{t}{t_*} = \frac{1}{2}\ln\frac{u'^2+(r-1)^2}{4}. \quad (4.1)$$

Here, we have used the asymptotic form for the Legendre function ($Q_{|m|}(\coth\xi) \approx \xi$ for $\xi \longrightarrow \infty$), which can be obtained from the hypergeometric representations of this function (see [7]). Expressions (2.4) and (4.1) lead to the following expression for the coordinate:

$$\frac{z}{c_0 t_*} = -\frac{|m|}{m}\arctan\frac{u'}{r-1}.$$

Thus, we obtain the formulas

$$\rho_* = r^{2m} = 1-4m\exp\frac{t}{t_*}\cos\frac{z}{c_0 t_*}, \qquad \frac{v}{c_0} = 4m\exp\frac{t}{t_*}\sin\frac{z}{c_0 t_*} \quad (4.2)$$

for the density and velocity, which describe perturbations periodic in coordinate z.

Let us consider the general case in which the Green function G can be expressed by formula (3.3). Using formulas (2.4) and (3.3), we obtain

$$z' = \pm e_* B(\zeta,\zeta_0)\sum_{n=0}^\infty g_n F_n, \quad (4.3)$$

where the plus and minus signs are used for $y - y_0 > 0$ and $y - y_0 < 0$, respectively, and g_n is the nth term of sum (3.3) for G,

$$F_n = \frac{1}{q_n}\left\{1+\frac{\zeta}{m}\frac{d}{d\zeta}\ln[B(\zeta,\zeta_0)L_n^{|m|}(\zeta)]\right\}. \quad (4.4)$$

It should be noted that factor ζ/m in the formula for F_n contains azimuthal number m itself and not its absolute value. The nonrelativistic limit of formula (4.3) can be obtained in the same way as in a transition from expression (3.3) to (3.4).

Let us consider in detail this transition for a Chaplygin gas, for which dependences $\rho_*(z,t)$ and $v(z,t)$ in the nonrelativistic case are described by explicit formulas. For quasi-Chaplygin media with $m < 0$, including the Chaplygin gas, formula (4.3) for $v/c \longrightarrow 0$ leads to the formula

$$\frac{z'}{c_0 t_*} = \pm\pi r^{|m|+1}\int_0^\infty J_{|m|+1}(qr)J_{|m|}(q)\times\exp(-|u'|q)dq, \quad (4.5)$$

in which the plus sign is take for $u' = v/2|m|c_0 < 0$ and the minus sign corresponds to $u' > 0$. We will consider below only a Chaplygin gas for which $m = -1/2$. For such a gas, integral (4.5) can easily be evaluated since the Bessel functions appearing in it have a simple form:

$$J_{1/2}(q) = \sqrt{\frac{2}{\pi q}}\sin q,$$

$$J_{3/2}(q) = \sqrt{\frac{2}{\pi q}}\left(\frac{1}{q}\sin q - \cos q\right).$$

Integrating expression (4.5) and considering that formula (3.6) assumes in the present case the form (see [7])

$$\frac{t}{t_*} = \frac{1}{2}\ln\left[\frac{u'^2 + (r-1)^2}{u'^2 + (r+1)^2}\right], \tag{4.6}$$

we obtain, using the Lorentz transformation, the following expression for coordinate z pertaining to the laboratory system of coordinates:

$$\frac{z}{c_0 t_*} = -\arctan\frac{2u'r}{u'^2 + 1 - r^2}. \tag{4.7}$$

Inverting formulas (4.6) and (4.7), we obtain the explicit dependences for the density and velocity on the coordinate and time,

$$\rho_*(z,t) = \frac{1}{r} = \frac{-\sinh(t/t_*)}{\cosh(t/t_*) - \cos(z/c_0 t_*)}, \qquad \frac{v(z,t)}{c_0} = \frac{\sin(z/c_0 t_*)}{\sinh(t/t_*)}, \tag{4.8}$$

which were given in [1, 2]. For $t \longrightarrow -\infty$, these expressions are transformed into expressions (4.2) for $m = -1/2$, which describe the linear stage of evolution of a perturbation.

5. UNSTABLE MEDIA DESCRIBED BY RELATIVISTIC QUASI-CHAPLYGIN EQUATIONS

We will consider below the following examples: a 1D Chaplygin gas, a Van der Waals gas in the instability domain, and a cylinder of a liquid with surface tension. In the nonrelativistic case, the evolution of these media can be described by quasi-Chaplygin equations (1.1) and (1.2), in which the azimuthal numbers are $m = -1/2$, 1, and –2, respectively (see [1, 2]). We will prove that the dynamics of these quasi-Chaplygin (or quasi-gaseous) media in the relativistic case can be described by RQCEs (1.3) and (1.4). To this end, we use the equation of motion (1.5) and continuity equation (1.6) in relativistic fluid dynamics. Disregarding pressure in the formula for entropy ($p \ll \rho c^2$), we can write these equations for the 1D case,

$$\left(\gamma\frac{\partial}{\partial\tau} + u\frac{\partial}{\partial z}\right)y = -\frac{p_0}{\rho_0 c^2}\frac{\rho_0}{\rho}\left(u\frac{\partial}{\partial\tau} + \gamma\frac{\partial}{\partial z}\right)\frac{p}{p_0}, \tag{5.1}$$

$$\frac{\partial}{\partial\tau}(n\gamma) + \frac{\partial}{\partial z}(nu) = 0, \tag{5.2}$$

where all quantities have the previous meaning and the zero subscript corresponds to unperturbed values of the corresponding quantities. The continuity equation (5.2) can be written in the form (1.4) for $\rho_* = n$.

One-dimensional Chaplygin gas. This hypothetic gas considered by Chaplygin [9] is characterized by a peculiar adiabat $p = p_0\rho_0/\rho$ along which the pressure increases upon a decrease in density. If we introduce the dimensionless effective density $\rho_* = \rho/\rho_0$, Eqs. (5.1) and (5.2) for such a gas assume the form of the RQCEs (1.3) and (1.4), in which $m = -1/2$ and $c_0^2 = p_0/\rho_0$.

Van der Waals gas in an unstable domain. The equation of state for real gases can be approximately described by the familiar Van der Waals model

$$p = 8T\rho(3-\rho)^{-1} - 3\rho^2,$$

where all quantities are reduced to their values at the critical point [10]. We will consider an isothermal process. It is well known that at a temperature lower than the critical point ($T < 1$), region $0 \le \rho \le 3$ contains an interval of values of ρ in which $dp/d\rho < 0$, while $dp/d\rho \ge 0$ to the left and right of this region. Standing perturbations increasing with time emerge in the gas precisely in this interval of densities, where the derivative $dp/d\rho$ is smaller than zero. The evolution of such perturbations can be described analytically using quasi-Chaplygin equations. For this purpose, we must represent $dp/d\rho$ as a power function of density. It can easily be verified that this can be done only at low temperatures in the interval $8T/9 \ll \rho \ll 3$, in which $dp/d\rho \approx -6\rho < 0$; this leads to RQCEs with an azimuthal number of $m = 1$. However, the gas pressure is negative in this density range and, hence, the quasi-Chaplygin description is formal in the present case.

A cylinder of liquid with surface tension. Let us consider perturbations of such a cylinder, which are distributed along its axis and which split the cylinder into drops due to surface tension. For the long-wave perturbations that will be considered below, the pressure produced by surface tension can be described by the formula $p = \sigma_*/a(z,t)$, where σ_* is the surface tension and $a(z,t)$ is the radius of the cylinder. We will use for the same perturbations the narrow channel or jet approximation [6], in which the pressure, density and

longitudinal velocity of a liquid in a cylinder are assumed to be constant over the cross section. Under these conditions, the equation of longitudinal motion assumes the form (5.1), while the continuity equation, as shown in [1, 4], has the form of Eq. (5.2), in which number density n should be replaced by the effective density $\rho_* = \rho a^2/\rho_0 a_0^2$. Further, we assume that the density of the liquid changes insignificantly (i.e., we set ρ = const). As a result, the equation of motion and the continuity equation can be written in the form of the RQCEs, in which $m = -2$, $c_0^2 = \sigma_*/2\rho_0 a_0$, and $\rho_* = a^2/a_0^2$.

This instability leads to extrusion of particles from waists to bulge regions located between waists; liquid particles are accelerated in this process. At the instant the waists rupture, the energy distribution function for the macroscopic motion of particles is completely formed. We denote this function by $F(E)$, where $E = \gamma Mc^2$ is the energy and M is the rest mass of a particle. Let us calculate the function $F(E)$. An analogous problem was considered in [4] for a skinned plasma pinch. Here, we repeat the calculation procedure as applied to our problem. The number of particles over an element of length dz of a cylinder of radius $a(z, t)$ is $dN = \gamma n \pi a^2 dz$, where γn and n are the number densities of particles in the laboratory and intrinsic frames of reference. Consequently, we obtain the equation

$$F(E) = \frac{dN}{dE} = \pi a_0^2 n_0 \rho_* \frac{\gamma}{Mc^2 u}\frac{dz}{dy} \tag{5.3}$$

for function $F(E)$, in which zero indices correspond to unperturbed values of quantities. Evaluating the derivative

$$\frac{dz}{dy} = \frac{\partial z}{\partial y} + \frac{\partial z}{\partial \zeta}\frac{d\zeta}{dy},$$

we must bear in mind that the $F(E)$ spectrum is calculated at a fixed instant $\tau(\zeta, y)$ = const, which gives

$$\frac{d\zeta}{dy} = -\frac{\partial \tau}{\partial y}\left(\frac{\partial \tau}{\partial \zeta}\right)^{-1}.$$

As a result, we obtain

$$F(E) = K\rho_*\frac{\gamma}{u}\left(\frac{\partial z}{\partial y}\frac{\partial \tau}{\partial \zeta} - \frac{\partial z}{\partial \zeta}\frac{\partial \tau}{\partial y}\right)\left(\frac{\partial \tau}{\partial \zeta}\right)^{-1}, \tag{5.4}$$

where

$$K = \frac{\pi \alpha_0^2 n_0}{Mc^2} = \text{const}.$$

In these formulas, τ and z are the time and coordinate in the laboratory system of coordinates, which are connected with time τ' and coordinate z' in the intrinsic system of coordinates via the Lorentz transformation (2.3).

Henceforth, we will consider only perturbations that are periodic along the cylinder. In accordance with the general theory of QCEs [1, 2], such perturbations can be described by the "Coulomb" solution to Eq. (2.7), i.e., by the Green function (3.3), in which $\zeta_0 = \varepsilon^2|m|$ and $y_0 = 0$. At the instant of rupture of waists, the radius of bulges between the waists tends to infinity so that $\zeta \longrightarrow 0$ in the bulges. In this case, for $m = -2$, we obtain from Eqs. (3.3) and (2.4) the following expressions for the bulge region: $\tau' \approx \zeta^2 f_0(y)$ and $z' \sim \zeta^3$. In these expressions, we have

$$f_0(y) = -\frac{1}{4}\sum_0^\infty \frac{n! L_n^2(0) L_n^2(2\varepsilon^2)}{q_n \Gamma(n+3)} \exp(-|y| q_n), \qquad q_n = \sqrt{1 + \frac{1+n}{2}}, \tag{5.5}$$

where L_n^2 is a Laguerre polynomial with superscript 2. Using Lorentz transformations (2.3), we obtain from these formulas the following expressions for τ and z:

$$\tau \approx \zeta^2 f_0(y)\cosh y, \quad z \approx \zeta^2 f_0(y)\sinh y.$$

Substituting these formulas into Eq. (5.4) and considering that $\rho_* = 4\varepsilon^4/\zeta^2$, we obtain the following expression for the $F(E)$ spectrum:

$$F(E) = \text{const}\frac{f_0(y)}{\sinh y}. \tag{5.6}$$

For ultrarelativistic energies, we retain in formula (5.5) for $f_0(y)$ only the term with $n = 0$. Considering that $E = \gamma Mc^2 \sim e^y$ for $\gamma \gg 1$, we find that the spectrum is given by

$$F(E \gg Mc^2) \sim E^{-(1+\sqrt{3/2})}. \tag{5.7}$$

Plasma pinch. Concluding the section, we recall that the relativistic dynamics of skinned plasma pinch in zero longitudinal magnetic field can also be described by the RQCEs with azimuthal number $m = -1$ (see [4]).

6. CONCLUSIONS

Quasi-Chaplygin equations describe the evolution of many unstable media (with a negative compressibility) encountered in nature. Analytically, these media differ in the value of parameter m known as the azimuthal number. A distinguishing feature of these nonlinear equations is that they have analytic solutions. In view of these two circumstances, it is natural to carry out a relativistic generalization of quasi-Chaplygin

equations and construct analytic solutions for them. This was done in the present study, In this generalization, we proceeded from the equations of one-liquid relativistic fluid dynamics as well as from the results obtained in [4], where a relativistic generalization of quasi-Chaplygin equations for $m = -1$ was obtained for a plasma pinch. In [4], a relativistic pinch was treated as a source of acceleration for cosmic rays.

In our work, we carry out this generalization for equations with an arbitrary value of azimuthal number m and give examples of the media (Chaplygin gas, Van der Waals gas in an unstable domain, and a cylinder of liquid with surface tension) whose dynamics can be described by the proposed relativistic quasi-Chaplygin equations. An analytic solution to these nonlinear equation is obtained for the 1D case.

ACKNOWLEDGMENTS

The author is grateful to Prof. B.A. Trubnikov for a detailed discussion of this work.

REFERENCES

1. S. K. Zhdanov and B. A. Trubnikov, *Quasi-Gaseous Unstable Media* (Nauka, Moscow, 1991) [in Russian].
2. B. A. Trubnikov, S. K. Zhdanov, and S. M. Zverev, *Hydrodynamics of Unstable Media* (CRC Press, Boca Raton, 1996).
3. A. V. Aksenov, Doctoral Dissertation in Physics and Mathematics (Moscow State Univ., Moscow, 2004).
4. V. P. Vlasov, S. K. Zhdanov, and B. A. Trubnikov, Pis'ma Zh. Éksp. Teor. Fiz. **49**, 581 (1989) [JETP Lett. **49**, 667 (1989)].
5. B. A. Trubnikov, Vopr. At. Nauki Tekh. (Ukr.) **2**, 78 (2005); Zh. Éksp. Teor. Fiz. **128**, 183 (2005).
6. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics*, 4th ed. (Nauka, Moscow, 1988; Pergamon, New York, 1987).
7. *Handbook of Mathematical Functions*, Ed. by M. Abramowitz and I. A. Stegun (Dover, New York, 1971; Nauka, Moscow, 1979).
8. I. S. Gradshteĭn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Fizmatlit, Moscow, 1963; Academic, New York, 1980).
9. S. A. Chaplygin, *Selected Works* (Nauka, Moscow, 1976) [in Russian].
10. M. A. Leontovich, *Introduction to the Thermodynamics. Statistical Physics* (Nauka, Moscow, 1983) [in Russian].

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 101, No. 6, 2005, pp. 979–988.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 6, 2005, pp. 1123–1133.

ATOMS, MOLECULES, OPTICS

Soliton Generation via Stimulated Brillouin Self-Scattering under Slow Light Conditions

S. V. Sazonov
Kaliningrad State University, Kaliningrad, 236041 Russia
e-mail: barab@newmail.ru
Received February 11, 2005

Abstract—Acousto-optic soliton generation via stimulated Brillouin self-scattering is predicted for light propagating at the speed of sound under electromagnetically induced transparency conditions. As in stimulated Raman self-scattering, the frequency of the electromagnetic component is gradually Stokes shifted as its intensity increases; the acoustic component has no carrier frequency. This phenomenon is explained by the possibility of forward stimulated Brillouin scattering, which is forbidden in nondispersive media. In contrast to stimulated Raman self-scattering, the Stokes shift of the electromagnetic component approaches a constant limit after the pulse has propagated to a certain distance. It is shown that the predicted soliton generation does not involve any threshold condition and can occur at extremely low input pulse intensities. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Investigation of various acousto-optical effects is a major area of research in nonlinear optics. The most outstanding examples are stimulated Raman scattering (SRS) and stimulated Brillouin scattering (SBS) [1], in which scattered Stokes waves are generated by interaction of light with optical and acoustic phonons, respectively [2]. In both phenomena, an important role is played by energy transfer from the pump and Stokes waves to vibrational modes of the medium.

The discovery of stimulated Raman self-scattering (SRSS) of femtosecond optical pulses [3] has opened new prospects for SRS studies and applications, particularly in fiber optics [1]. The occurrence of this phenomenon does not depend on any threshold condition, because femtosecond pulses have wide bandwidths and therefore become coherently coupled with optical vibrational modes as soon as they enter the nonlinear medium [3, 4]. As the pulse continues to propagate in a dispersive medium, an increasingly Stokes-shifted soliton is generated. In a nondispersive medium, SRSS does not lead to soliton generation and the optical-pulse spectrum substantially broadens, splitting into lines [4].

Unlike SRS or SRSS, forward SBS is forbidden by energy and momentum conservation in elementary photon–phonon scattering events [1, 2]. This is true for weakly dispersive media, where the relative SBS frequency shift can barely reach 0.01%. Therefore, the refractive indices corresponding to the input and scattered-wave frequencies cannot be substantially different, and light scattering by acoustic phonons cannot lead to any phenomenon analogous to soliton generation via SRSS.

However, recent experimental observations of ultraslow light propagation in gases [5] and solids [6] under electromagnetically induced transparency (EIT) conditions have revealed new possibilities for controlling dispersion characteristics of optical materials. It was shown in [7] that SBS characteristics drastically change as the group velocity of light approaches the speed of sound in the medium. The most important consequence is the possibility of forward scattering due to combined effects of nonlinearity and dispersion [7]. Since solitons can be generated in a nonlinear dispersive medium, a question arises about the possibility of stimulated Brillouin self-scattering (SBSS, an analog of SRSS) leading to simultaneous generation of a Stokes-shifted soliton and a coherent acoustic phonon. This possibility is explored in the present study.

The paper is organized as follows. In Section 2, nonlinear wave equations are derived for collinearly propagating optical and longitudinal acoustic waves under EIT conditions. The slowly-varying-envelope approximation is applied to rewrite the general system of integrodifferential equations as Zakharov-type equations for an optical pulse envelope, which are then reduced to the Yajima–Oikawa system in a unidirectional approximation. In Section 3, the physics of the soliton solution to this system is analyzed. The results of this analysis and numerical estimates suggest that soliton generation via SBSS can be observed in experiments. In Section 4, the defocusing of an SBSS acousto-optic soliton by transverse perturbations is analyzed. The Conclusions section summarizes the principal results and highlights the similarities and distinctions between SBSS and SRSS.

1063-7761/05/10106-0979$26.00

2. NONLINEAR WAVE EQUATIONS

Consider an isotropic medium characterized by an ultraslow group velocity v_g of light propagation under EIT conditions, which are characterized by an almost vanishing resonant absorption coefficient and coherent population trapping [8].

Following [7], let us start with the energy and momentum conservation laws for an SBS event:

$$\omega_{in} = \omega_{out} + \Omega, \quad \mathbf{k}_{in} = \mathbf{k}_{out} + \mathbf{k}_a, \tag{1}$$

where ω_{in} and ω_{out} are the frequencies of the incident and scattered photons, respectively; Ω is the acoustic-phonon frequency; and $\mathbf{k}_{in}$, $\mathbf{k}_{out}$, and $\mathbf{k}_a$ are the corresponding wavevectors.

The second equation in (1) yields

$$k_a^2 = (k_{in} - k_{out})^2 + 4k_{in}k_{out}\sin^2\frac{\theta}{2}, \tag{2}$$

where θ is the angle between $\mathbf{k}_{in}$ and $\mathbf{k}_{out}$. Since $k_a = \Omega/a$ (a is the speed of sound in the medium) and $k_{in,\,out} = \omega_{in,\,out}n_{in,\,out}/c \gg k_a$ (c is the speed of light in free space, and $n_{in,\,out} = n(\omega_{in,\,out})$ are the refractive indices at ω_{in} and ω_{out}, respectively), it holds that

$$k_{in} - k_{out} = \frac{\partial k}{\partial \omega}(\omega_{in} - \omega_{out}) = \frac{\Omega}{v_g},$$

where v_g is the group velocity of light at ω_{in}. Therefore, Eq. (2) can be rewritten as

$$\left(\frac{1}{a^2} - \frac{1}{v_g^2}\right)\Omega^2 = 4\frac{n_{in}\omega_{in}}{c}\left(\frac{n_{in}\omega_{in}}{c} - \frac{\Omega}{v_g}\right)\sin^2\frac{\theta}{2}. \tag{3}$$

In the general case, Eq. (3) should be treated as a quadratic equation for the Stokes shift Ω due to SBS. For a nondispersive medium, $v_g \approx c/n_{in}$. Since $n_{in} \sim 1$, $c \gg a$, and $\Omega \ll \omega_{in}$, the second terms in both parentheses can be neglected and Eq. (3) reduces to a well-known expression [2, 9]:

$$\Omega = 2\frac{n_{in}\omega_{in}a}{c}\left|\sin\frac{\theta}{2}\right|.$$

If $v_g = a$ under EIT conditions, then it follows from Eq. (3) that $\Omega = n_{in}\omega_{in}a/c$ for $\theta \neq 0$. In the case of $\theta = 0$ (forward scattering), Eq. (3) is an identity for any Ω; i.e., the Stokes shift due to forward SBS cannot be determined from energy and momentum conservation in a scattering event.

The polarization response $P_r^{(0)}(\mathbf{r}, t)$ of an undeformed dispersive medium to a linearly polarized field $E(\mathbf{r}, t)$ can be represented as

$$P_r^{(0)}(\mathbf{r}, t) = \chi_m^{(0)}E(\mathbf{r}, t) + \int_0^\infty \chi_0(\tau)E(\mathbf{r}, t-\tau)d\tau, \tag{4}$$

where $\chi_m^{(0)}$ is the dielectric susceptibility of the host material (treated as a nondispersive medium in the case of an off-resonance optical-pulse carrier frequency) and $\chi_0(\tau)$ is the susceptibility of the dispersive medium consisting of resonant impurities (which are responsible for EIT).

The modulation of both host and resonant-impurity susceptibilities due to the coupling between electromagnetic waves and acoustic phonons is described by performing the change (see [2, 9])

$$\chi_m^{(0)} \longrightarrow \chi_m = \chi_m^{(0)} - \gamma_1 u, \tag{5}$$

where $u(\mathbf{r}, t)$ is longitudinal strain, $\gamma_1 = -(\partial\chi_m/\partial u)_0$ is the electrostriction coefficient [9], and the index "0" denotes the properties of the undeformed medium.

Modulation of the nonlocal susceptibility $\chi_0(\tau)$ in expression (4) is taken into account by substituting $\chi_0(t)$ with $\chi(\tau, u(\mathbf{r}, t))$. The delay in the impurity polarization response to strain is neglected here, because the acoustic phonon spectrum is separated from the resonant optical absorption spectrum.

The Taylor series expansion of $\chi(\tau, u(\mathbf{r}, t))$ in $u(\mathbf{r}, t)$ about τ yields

$$\chi(\tau, u(\mathbf{r}, t)) = \chi_0(\tau) - \Gamma(\tau)u(\mathbf{r}, t), \tag{6}$$

where

$$\Gamma = -\left(\frac{\partial\chi}{\partial u}\right)_0.$$

Expressions (4)–(6) are used to write the photon–phonon coupling term responsible for SBS as

$$V_{int} = \frac{1}{2}u\left[\gamma_1 E^2 + E\int_0^\infty \Gamma(\tau)E(\mathbf{r}, t-\tau)d\tau\right]. \tag{7}$$

The polarization response modulated by the coupling with acoustic phonons is

$$P_r(\mathbf{r}, t) = P_r^{(0)}(\mathbf{r}, t) - u\left[\gamma_1 E + \int_0^\infty \Gamma(\tau)E(\mathbf{r}, t-\tau)d\tau\right]. \tag{8}$$

Since the first terms in the brackets in (7) and (8) correspond to acoustic modulation of the susceptibility of a nonresonant host crystal, they do not represent the effect of SBS on EIT. The second terms in these brackets describe the effect of acoustic phonons on the nonlocal polarization response of a resonant impurity, which is responsible for slow light propagation under EIT conditions. In other words, these terms describe the effect of SBS on EIT conditions. As mentioned above,

the effect of EIT on SBS primarily manifests itself by a decrease in the group velocity of light to the speed of sound and the ensuing possibility of forward scattering.

Expressions (4), (7), and (8) are combined with Maxwell's equation for the electric field of the optical pulse,

$$\Delta E - \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2} = \frac{4\pi}{c^2}\frac{\partial^2 P_r}{\partial t^2}, \tag{9}$$

and the Hamiltonian density of the acoustic field,

$$V_a = \frac{p^2}{2\rho_m} + \frac{1}{2}\rho_m a^2 (\nabla s)^2, \tag{10}$$

where ρ_m is the mean density of the medium, a is the longitudinal speed of sound, s is the displacement field related to the strain field u by the equation $u = \partial s/\partial z$ (both optical and acoustic pulses propagate along the z axis), and p is the displacement momentum density.

Using the Hamiltonian equations

$$\frac{\partial s}{\partial t} = \frac{\delta H}{\delta p}, \quad \frac{\partial p}{\partial t} = -\frac{\delta H}{\delta s},$$

with

$$H = \int (V_a + V_{int}) d^3\mathbf{r},$$

and substituting (8) into (9), we obtain

$$\Delta E - \frac{n_m^2}{c^2}\frac{\partial^2 E}{\partial t^2} = \frac{4\pi}{c^2} \times \frac{\partial^2}{\partial t^2}\left\{\int_0^\infty \chi(\tau) E(\mathbf{r}, t-\tau) d\tau - u\left[\gamma_1 E + \int_0^\infty \Gamma(\tau) E(\mathbf{r}, t-\tau) d\tau\right]\right\}, \tag{11}$$

$$\Delta u - \frac{1}{a^2}\frac{\partial^2 u}{\partial t^2} = -\frac{1}{2\rho_m a^2} \times \frac{\partial^2}{\partial z^2}\left[\gamma_1 E^2 + E\int_0^\infty \Gamma(\tau) E(\mathbf{r}, t-\tau) d\tau\right], \tag{12}$$

where

$$n_m = \sqrt{1 + 4\pi\chi_m^{(0)}}$$

is the refractive index of the host crystal.

Integrodifferential equations (11) and (12) provide a self-consistent model of SBS in a dispersive medium.

The electric field of a quasi-monochromatic pulse with carrier frequency ω and wavenumber k can be represented as

$$E(\mathbf{r}, t) = \psi(\mathbf{r}, t)\exp[i(\omega t - kz)] + \text{c.c.}, \tag{13}$$

where $\psi(\mathbf{r}, t)$ is a slowly varying envelope (see [10]).

Expression (13) is substituted into (11), and the function $\psi(\mathbf{r}, t-\tau)$ is represented by the Taylor series expansion in τ. Restricting the expansion to the lowest order dispersive terms, we write

$$\int_0^\infty \chi(\tau) E(\mathbf{r}, t-\tau) d\tau = \left[\chi(\omega)\psi - i\left(\frac{\partial\chi}{\partial\omega}\right)\frac{\partial\psi}{\partial t} - \frac{1}{2}\left(\frac{\partial^2\chi}{\partial\omega^2}\right)\frac{\partial^2\psi}{\partial t^2} + \frac{i}{6}\left(\frac{\partial^3\chi}{\partial\omega^3}\right)\frac{\partial^3\psi}{\partial t^3}\right]\exp[i(\omega t - kz)] + \text{c.c.}, \tag{14}$$

where

$$\chi(\omega) = \int_0^\infty \chi(\tau)\exp(-i\omega\tau) d\tau.$$

The integral

$$\int_0^\infty \Gamma(\tau) E(\mathbf{r}, t-\tau) d\tau$$

in (11) and (12) is represented by an analogous expansion.

Substituting (13) and (14) into Eqs. (12) and (11) and dropping the fastest oscillating terms, we obtain

$$i\left(\frac{\partial\psi}{\partial z} + \frac{1}{v_g}\frac{\partial\psi}{\partial t}\right) + \frac{k_2}{2}\frac{\partial^2\psi}{\partial t^2} - i\frac{k_3}{6}\frac{\partial^3\psi}{\partial t^3} = -\frac{2\pi\omega}{cn}u\left[\gamma\psi - i\left(\frac{\partial\gamma_2}{\partial\omega}\right)\frac{\partial\psi}{\partial t}\right] + \frac{c}{2n\omega}\Delta_\perp\psi, \tag{15}$$

$$\frac{\partial^2 u}{\partial z^2} - \frac{1}{a^2}\frac{\partial^2 u}{\partial t^2} = -\frac{1}{\rho_m a^2} \times \frac{\partial^2}{\partial z^2}\left[\gamma|\psi|^2 + \frac{i}{2}\left(\frac{\partial\gamma_2}{\partial\omega}\right)\left(\psi\frac{\partial\psi^*}{\partial t} - \psi^*\frac{\partial\psi}{\partial t}\right)\right] - \Delta_\perp u. \tag{16}$$

Here,

$$v_g = \frac{c}{n + \omega(\partial n/\partial\omega)}$$

is the group velocity of the optical pulse,

$$n = [1 + 4\pi(\chi_m^{(0)}) + \chi(\omega)]^{1/2}$$

is the total refractive index,

$$k_2 = \frac{\partial(1/v_g)}{\partial\omega}$$

is the group-velocity dispersion parameter,

$$k_3 = \frac{2\pi}{cn}\left[\frac{6}{\omega}\frac{\partial}{\partial\omega}\left(\omega\frac{\partial\chi}{\partial\omega}\right) + \omega\frac{\partial^3\chi}{\partial\omega^3}\right],$$

$$\gamma = \gamma_1 + \gamma_2,$$

$$\gamma_2(\omega) = \int_0^\infty \Gamma(\tau)\exp(-i\omega\tau)d\tau,$$

and $\Delta_\perp$ is the transverse Laplace operator.

Whereas electromagnetic field is represented by an envelope in Eqs. (15) and (16), acoustic waves are represented by strain, because the acoustic pulse has no carrier frequency in the general case.

Since $\chi(\omega) = 0$ [8] and $\partial^2 n/\partial\omega^2 = 0$ [11] at the absorption-line center frequency under EIT conditions, it follows that $n = n_m$ and $k_2 = 2(\partial n/\partial\omega)/c > 0$ (group dispersion is positive at the absorption-line center).

The resonant-impurity concentration in solids used in experiments on EIT varies within 0.05–0.1% of the concentration of atoms of the nonresonant host material [6]. Therefore, $\gamma_2 \ll \gamma_1$ and it can be assumed that $\gamma \approx \gamma_1$.

The ratios of the terms containing the factor $(\partial\gamma_2/\partial\omega)$ to the first terms in the respective brackets are on the order of $(\omega\tau_p)^{-1} \ll 1$, where τ_p is the pulse duration. The former terms characterize the nonlinear dispersion due to the effect of SBS on EIT. Both nonlinear dispersion and the last term on the left-hand side of Eq. (15) can be neglected for nanosecond optical pulses (with $\omega \sim 10^{15}$ s^{-1}), because $(\omega\tau_p)^{-1} \sim 10^{-6}$. Then, Eqs. (15) and (16) reduce to Zakharov-type equations [12] in the absence of transverse perturbations.

The analysis of system (15), (16) presented here shows that the effect of SBS on EIT conditions is negligible in the case of a quasi-monochromatic resonant optical pulse. However, EIT substantially modifies SBS characteristics; in particular, forward scattering of light with $v_g = a$ becomes possible. This equality is the condition for the Zakharov–Benney resonance between long and short waves [13]: the group velocity of a short-wave (electromagnetic) component is equal to the phase velocity of a long (acoustic) wave. Indeed, it follows from Eqs. (1) written for collinear propagation that

$$a = \frac{\Omega}{k_a} = \frac{\omega_{in} - \omega_{out}}{k_{in} - k_{out}} \approx \frac{\partial\omega}{\partial k} = v_g.$$

This condition corresponds to the highest efficiency of coupling between acoustic waves and ultraslow light. In this case, Eq. (16) can be conveniently rewritten in a quasi-unidirectional approximation [13]. Since (5) is actually an expansion, the first term on the right-hand side in (16) is small. In the paraxial approximation

$$\Delta_\perp u \ll \frac{\partial^2 u}{\partial z^2},$$

which is valid if $(l_\parallel/R)^2 \ll 1$ ($l_\parallel$ and R are the longitudinal and transverse pulse dimensions, respectively), transverse perturbations can be treated as long-wavelength ones; i.e., the acousto-optic pulse is quasi-one-dimensional. Introducing the "local" time

$$\tau = t - z/a = t - z/v_g$$

and the "slow" coordinate

$$\zeta = \mu z$$

with a small parameter μ on the order of the right-hand side in (16), neglecting terms on the order of μ^2, and dropping the terms representing nonlinear dispersion and third-order linear dispersion, we rewrite Eqs. (15) and (16) as

$$i\frac{\partial\psi}{\partial z} + \frac{k_2}{2}\frac{\partial^2\psi}{\partial\tau^2} = -\alpha u\psi + \frac{c}{2n_m\omega}\Delta_\perp\psi, \qquad (17)$$

$$\frac{\partial u}{\partial z} = \beta\frac{\partial}{\partial\tau}(|\psi|^2) + \frac{a}{2}\Delta_\perp\int_{-\infty}^{\tau} u d\tau', \qquad (18)$$

where

$$\alpha = \frac{2\pi\omega\gamma}{cn_m}, \quad \beta = \frac{\gamma}{2\rho_m a^3}.$$

3. ACOUSTO-OPTIC SOLITON

Equations (17) and (18) are equivalent to the Yajima–Oikawa system [14], which has the one-soliton solution

$$\psi = \psi_m\exp(-i(\Omega t - qz))\,\mathrm{sech}\left(\frac{t - z/v}{\tau_p}\right),$$
$$u = u_m\,\mathrm{sech}^2\left(\frac{t - z/v}{\tau_p}\right), \qquad (19)$$

where

$$\psi_m = \frac{|k_2|}{\tau_p}\sqrt{\frac{\Omega}{\alpha\beta}}, \quad u_m = \frac{k_2}{\alpha\tau_p^2},$$

$$q = \frac{\Omega}{a} + \frac{k_2}{2}\left(\frac{1}{\tau_p^2} - \Omega^2\right),$$

and the propagation velocity v is determined by the relation

$$\frac{1}{v} = \frac{1}{a} - k_2\Omega. \quad (20)$$

According to the expression for u_m, the acoustic component corresponds to compressive strain if $\gamma < 0$ and to dilatational strain otherwise.

Solution (19) contains two free parameters, which can be defined as the duration τ_p and amplitude ψ_m of the electromagnetic component. The parameter Ω should be interpreted as a "nonlinear" shift of the carrier frequency due to forward SBS. It follows from the expressions for α, β, and ψ_m that $\Omega \geq 0$. In view of (13) and (19), this implies that the optical pulse frequency is Stokes shifted,

$$\omega \longrightarrow \omega - \Omega;$$

i.e., energy is transferred from the photon to the phonon generated in each SBS event. As shown in Section 2, the value of Ω cannot be determined by using energy and momentum conservation laws for a photon–phonon scattering event if $v_g = a$ and $\theta = 0$.

The value of Ω depends on the input pulse parameters and can be determined by solving an appropriate inverse scattering problem [14]. Alternatively, the shift can be found by solving system (17), (18) numerically for different values of the input pulse parameters. Note also that the asymptotic value of the Stokes shift is proportional to the input pulse intensity $I_{opt} \sim \psi_m^2$, which characterizes an ensemble of photons rather than an individual photon. The shifted optical pulse is detuned from resonance with a quantum transition in impurity atoms, and its group velocity increases according to (20), because the slowest light propagation under EIT conditions is observed in the transparency window, which corresponds to exact resonance with the impurity transition [8].

It follows from the analysis presented above that there are no formal restrictions for the value of Ω in forward scattering when $v_g = a$. However, there exists an upper limit for Ω for physical reasons. First, $\Omega \ll \omega$ since ψ is a slowly varying envelope (see (13) and (19)). Second, the value of Ω must be much smaller than the EIT window, which is determined by the pump Rabi frequency Ω_p: $\Omega \ll \Omega_p$. Note that the entire optical-pulse spectrum must remain within the transparency window. This restriction can be written as

$$\tau_p \gg \frac{1}{\Omega_p}.$$

The present analysis of soliton solution (19) shows that stimulated Brillouin self-scattering analogous to SRSS can be implemented under conditions of forward SBS (light propagation at the speed of sound). Indeed, the Stokes shift Ω of the pulse carrier frequency gradually increases with the input pulse intensity to the upper limit indicated above, while there is no lower limit for Ω. Therefore, SBSS is analogous to SRSS in that it does not involve any threshold condition [3, 4].

It follows from Eqs. (15) and (16) (or (17) and (18)) that the input pulse generates an acoustic video pulse (a carrier-free elastic wave), whereas the acoustic pulse cannot generate an electromagnetic wave in the absence of incident light.

Soliton generation via SBSS can be interpreted as gradual energy transfer from an optical pulse to an elastic wave. Eventually, the shift approaches a limit value; i.e., the wave evolves into soliton (19) with a constant Ω. This "Stokes saturation" is explained by the lower efficiency of the photon–phonon coupling (responsible for forward SBS) in a pulse detuned from the center of the EIT window.

Let us estimate the parameters of an acousto-optic soliton described by (19). Experiments on ultraslow light propagation were performed at a temperature of 5 K on an insulator (Y_2SiO_5 crystal) doped with rare-earth (praseodymium) ions to a concentration of 0.05% [6], because these ions are characterized by narrow inhomogeneous broadening of transition lines. The group velocity of light with input intensity $I_p \approx 470$ W/cm^2 was slowed down to $v_g \approx 4.5 \times 10^3$ cm/s, while the speed of sound was higher by two orders of magnitude. Since the group velocity of ultraslow light under EIT conditions varies as does the pump Rabi frequency squared (i.e., linearly with I_p) [5], the condition $v_g \approx a$ is satisfied when $I_p \sim 50$ kW/cm^2. The corresponding Rabi frequency is

$$\Omega_p \approx \frac{d}{\hbar}\sqrt{\frac{4\pi n_m I_p}{c}},$$

where d is the dipole moment of the resonant-impurity transition. For $d \approx 5 \times 10^{-18}$ SGSE units and $n_m \approx 2$, it follows that $\Omega_p \sim 10^{11}$ s^{-1}. Setting $\Omega \sim 10^{10}$ s^{-1} and $\tau_p \sim 1$ ns in (19) in accordance with conditions imposed above and using the fact that $a \ll c$, we obtain

$$\frac{1}{v_g} = \frac{1}{c}\left(n + \omega\frac{\partial n}{\partial \omega}\right) \approx \frac{\omega}{c}\frac{\partial n}{\partial \omega} \approx \frac{1}{a}.$$

Since $\partial^2 n/\partial\omega^2 = 0$ for an optical pulse resonant with an impurity transition, we have

$$k_2 = \left[\frac{\partial}{\partial\omega}\left(\frac{1}{v_g}\right)\right] = \frac{2}{c}\left(\frac{\partial n}{\partial \omega}\right) \approx \frac{2}{\omega a}. \quad (21)$$

If $n_m \approx 2$, $\gamma \sim 1$ [9], $\rho_m \approx 2$ g/cm^3, $a \approx 5 \times 10^5$ cm/s, and $\omega \approx 3 \times 10^{15}$ s^{-1}, then the intensity of the soliton's elec-

tromagnetic component is

$$I_{\text{opt}} \approx \frac{c}{4\pi}\psi_{\text{m}}^2 \approx \frac{n_{\text{m}}a}{\pi^2}\frac{\Omega}{\omega}\frac{\rho c^2}{(\omega\tau_{\text{p}}\gamma)^2} \sim 10\ \text{W/cm}^2.$$

The corresponding values of the relative strain amplitude u_{m} and acoustic pulse intensity I_{s} are

$$u_{\text{m}} \approx \frac{cn_{\text{m}}}{\pi a\gamma(\omega\tau_{\text{p}})^2} \sim 10^{-8},$$

$$I_{\text{s}} \approx \rho_{\text{m}}a^3u_{\text{m}}^2/2 \sim 10^{-6}\ \text{W/cm}^2.$$

Thus, a soliton whose duration is a few nanoseconds has an extremely low intensity, which is primarily determined by its electromagnetic component.

It follows from Eqs. (14) and (15) that

$$\frac{1}{v} \approx \frac{1}{a}\left(1 - 2\frac{\Omega}{\omega}\right).$$

Using the estimates above, we obtain $2\Omega/\omega \sim 10^{-5}$. Thus, the propagation velocity of an acousto-optic soliton described by (19) differs from the linear speed of sound only by a negligible 10^{-3}%.

Note that the electromagnetic component of the pulse vanishes when $\Omega = 0$; i.e., the input pulse energy is entirely converted into the energy of an elastic-strain soliton. To find conditions for this phenomenon to occur, one must solve a boundary value problem for Eqs. (17) and (18), which requires a separate analysis.

4. EFFECTS OF TRANSVERSE PERTURBATIONS

Experimental observability of soliton (19) depends on its stability with respect to transverse perturbations. The Ritz–Whitham averaged-Lagrangian method is applied here to allow for transverse dynamics [17].

The Lagrangian density corresponding to system (17), (18) is

$$L = \frac{i}{2}\left(\psi\frac{\partial\psi^*}{\partial z} - \psi^*\frac{\partial\psi}{\partial z}\right)$$

$$+ \frac{k_2}{2}\left|\frac{\partial\psi}{\partial\tau}\right|^2 + \alpha|\psi|^2\frac{\partial U}{\partial\tau} - \frac{c}{2n_{\text{m}}\omega}|\nabla_\perp\psi|^2 \qquad (22)$$

$$+ \frac{\alpha}{2\beta}\left[\frac{\partial U}{\partial z}\frac{\partial U}{\partial\tau} - \frac{a}{2}(\nabla_\perp U)^2\right],$$

where U is related to strain by the equation

$$u = \frac{\partial U}{\partial\tau}.$$

In view of the remark about the quasi-one-dimensional dynamics of the acousto-optic pulse made at the end of Section 2, a trial solution is sought in the form of modified one-dimensional soliton (19):

$$\psi = |k_2|\sqrt{\frac{\Omega}{\alpha\beta}}\rho\exp\left[-i\Omega\left(t - \frac{z}{a}\right) - \frac{n_{\text{m}}\omega}{c}\Phi\right]$$

$$\times \operatorname{sech}\left[\rho\left(t - \frac{z}{v}\right)\right], \qquad (23)$$

$$U = \frac{k_2}{\alpha}\rho\tanh\left[\rho\left(t - \frac{z}{v}\right)\right].$$

The dynamic variables ρ and Φ introduced here can be interpreted as the inverse soliton duration and the eikonal of its electromagnetic component, respectively. They are to be determined as functions of coordinates, while Ω is treated as a constant parameter of a well-developed soliton.

After substituting (23) into (22), an averaged Lagrangian is found by integration with respect to the "fast" variable τ (as done in [18] for different nonlinear equations):

$$\Lambda \equiv \frac{c\alpha\beta}{2n_{\text{m}}\omega\Omega k_2^2}\int_{-\infty}^{\infty} L\,d\tau = -\rho\frac{\partial\Phi}{\partial z} - \frac{1}{2}\rho(\nabla_\perp\Phi)^2$$

$$+ \frac{ck_2}{2n_{\text{m}}\omega}(\Omega^2\rho - \rho^3/3) - b\frac{(\nabla_\perp\rho)^2}{4\rho}, \qquad (24)$$

where

$$b = \frac{c}{3n_{\text{m}}\omega}\left[\left(\frac{\pi^2}{6} + 2\right)\frac{c}{n_{\text{m}}\omega} + \frac{a}{2\Omega}\right].$$

The Euler–Lagrange equations for ρ and Φ corresponding to (24) are

$$\frac{\partial\rho}{\partial z} + \nabla_\perp(\rho\nabla_\perp\Phi) = 0,$$

$$\frac{\partial\Phi}{\partial z} + \frac{(\nabla_\perp\Phi)^2}{2} + \frac{ck_2}{2n_{\text{m}}\omega}(\rho^2 - \Omega^2) = b\frac{\Delta_\perp\sqrt{\rho}}{\sqrt{\rho}}. \qquad (25)$$

System (25) has a one-dimensional solution equivalent to (19),

$$\rho = \frac{1}{\tau_{\text{p}}} = \text{const},$$

$$\Phi = \Phi_0(z) = \frac{ck_2}{2n_{\text{m}}\omega}\left(\Omega^2 - \frac{1}{\tau_{\text{p}}^2}\right)z,$$

which substantiates the validity of the averaged-Lagrangian method.

Transverse perturbations are taken into account in the left- and right-hand sides of the second equation in (25). According to the definition of b, this coefficient vanishes as $\omega \sim 1/\lambda \longrightarrow \infty$, which corresponds to the eikonal approximation in geometric acousto-optics (λ is the wavelength of the electromagnetic component). Therefore, the term $b\Delta_\perp\sqrt{\rho}/\sqrt{\rho}$ in (25) represents diffraction effects in the transverse pulse dynamics, while the term $(\nabla_\perp\Phi)^2/2$ is responsible for the transverse dynamics in the eikonal approximation (nonlinear refraction) [18, 19].

In the general case, nonlinear system (25) is difficult to analyze. As a first step, consider soliton (19) weakly perturbed by transverse perturbations:

$$\rho = \frac{1}{\tau_p} + \rho_1, \quad \Phi = \Phi_0 + \Phi_1,$$

where

$$\rho_1 \ll \frac{1}{\tau_p}, \quad \Phi_1 \ll \Phi_0.$$

Substituting

$$\rho_1, \Phi_1 \approx \exp[i(q_\parallel z + \mathbf{q}_\perp \cdot \mathbf{r}_\perp)],$$

into Eq. (25) linearized with respect to ρ_1 and Φ_1, we obtain the "dispersion" relation

$$q_\parallel^2 = \frac{1}{2}\left(\frac{ck_2}{n_m\omega\tau_p^2} + bq_\perp^2\right)q_\perp^2. \tag{26}$$

Since $\kappa_2, \Omega > 0$, this relation holds only if $q_\parallel$ is real. Therefore, acousto-optic soliton (19) is stable with respect to small transverse perturbations.

Before proceeding to a quantitative analysis of nonlinear behavior of transverse perturbations, note that system (25) with zero right-hand sides is formally equivalent to the continuity equation and Cauchy theorem for inviscid flows [20], with ρ, Φ, and z corresponding to fluid density, velocity potential, and time, respectively. Comparing the second equation in (25) with the Cauchy theorem written as

$$\frac{\partial\Phi}{\partial z} + \frac{(\nabla_\perp\Phi)^2}{2} + \int\frac{dP}{\rho} = \text{const},$$

where P corresponds to static pressure, we obtain the following "isentropic" equation

$$\frac{dP}{d\rho} = \frac{ck_2}{n_m\omega}\rho^2.$$

Then, the criterion for the soliton's stability with respect to self-focusing is identical to the stability condition of the "inviscid flow" described by Eqs. (25) with zero right-hand sides:

$$\frac{dP}{d\rho} > 0$$

(see [21, 22]). Since $k_2 > 0$ for the process considered here, we conclude that nonlinearity has a defocusing effect in the eikonal approximation. This is an expected result, because solitons can be generated under conditions of defocusing nonlinearity only in the spectral region of normal group dispersion [1, 10].

To perform a general analysis of transverse dynamics (including self-diffraction effects), we drop the assumption that transverse perturbations are small. Then, system (25) is equivalent to a quintic nonlinear Schrödinger equation:

$$i\frac{\partial Q}{\partial z} = -g\Delta_\perp Q + \frac{\eta}{g}|Q|^4 Q, \tag{27}$$

where

$$g = \pm\sqrt{\frac{b}{2}}, \quad \eta = \frac{ck_2}{4n_m\omega},$$

and the complex-valued function Q is defined by the relation

$$Q = \sqrt{\rho}\exp\left[\frac{i}{2g}\left(\frac{ck_2Q^2}{n_m\omega}z + \Phi\right)\right]. \tag{28}$$

Thus, stability analysis of the soliton described by Eq. (19) is equivalent to stability analysis of the spatial "beam" described by Eq. (27). Here, the integral moment method developed in [23] is applied.

Equation (27) corresponds to the "Hamiltonian"

$$\tilde{H} = \int\left(g^2|\nabla_\perp Q|^2 + \frac{\eta}{3}|Q|^6\right)dS \tag{29}$$

and entails the conservation of

$$N = \int|Q|^2 dS,$$

where the integral is performed over the xy plane. If the soliton radius squared is defined as the second-order moment

$$R^2 = \frac{1}{N}\int r^2|Q|^2 dS, \tag{30}$$

where r is radius in a cylindrical coordinate system

(see [23]), then it follows from Eq. (27) that

$$\frac{dR^2}{dz} = \frac{2}{N}\int \mathbf{j}\cdot\mathbf{r}\,dS, \tag{31}$$

where

$$\mathbf{j} = ig[Q(\nabla_\perp Q^*) - Q^*(\nabla_\perp Q)] = \rho\nabla_\perp\varphi.$$

The derivative of Eq. (31) yields

$$\frac{d^2R^2}{dz^2} = \frac{16}{N}\tilde{H} = \text{const}. \tag{32}$$

Since $k_2 > 0$, both η and $\tilde{H}$ are strictly positive; i.e., (32) implies that

$$\frac{d^2R^2}{dz^2} > 0.$$

Therefore, acousto-optic soliton (19) must exhibit defocusing behavior.

Assuming that the electromagnetic component has a plane wavefront at $z = 0$,

$$\nabla_\perp\varphi_{z=0} = 0,$$

and substituting the expression for $\mathbf{j}$ into (31), we obtain

$$\left(\frac{dR^2}{dz}\right)_{z=0} = 0.$$

Then, the solution to Eq. (32) is

$$R = R_0\sqrt{1 + (z/l)^2}, \tag{33}$$

where R_0 is the "initial" soliton radius and

$$l = R_0\sqrt{N/8\tilde{H}}$$

is the soliton defocusing length.

The first and second terms on the right-hand sides in (27) and (29) represent the contributions to transverse dynamics due to diffraction and nonlinear dispersion, respectively. These terms play a dominant role in the eikonal approximation and in the limit of $g \longrightarrow \infty$, respectively. Their relative importance can be quantified by introducing the dimensionless parameter

$$\delta \sim (R/\sqrt{ac}\tau_p)^2.$$

If $\tau_p \sim 1$ ns (see above) and $R \sim 30$ μm, then $\delta \sim 10^{-3}$, and the inequality $(l_\parallel/R)^2 \ll 1$ holds because $l_\parallel \sim a\tau_p \sim 10$ μm. Under these conditions, diffraction effects are much stronger than those due to nonlinear refraction; i.e., the first terms on the right-hand sides in (27) and (29) are much larger than the second ones. Since $|Q|^2 = \rho$, we have

$$|\nabla_\perp Q|^2 \sim \frac{\rho}{2R_0^2}.$$

Then,

$$l = l_d = \frac{R_0^2}{\sqrt{2b}},$$

where l_d is the diffractive spread. In this limit, Eq. (27) reduces to a linear equation. Suppose that a well-developed soliton has an axially symmetric Gaussian transverse profile with characteristic radius R_0 at $z = 0$:

$$Q(r, z = 0) = \sqrt{\rho(r, z = 0)} = \sqrt{\rho_0}\exp\left(-\frac{r^2}{2R_0^2}\right),$$

where ρ_0 is the inverse duration of the soliton on its centerline and r is radius in a cylindrical coordinate system. Then, the solution to Eq. (27) with zero last term is

$$Q(r, z) = \frac{\sqrt{\rho_0}}{1 + iz/l_d}\exp\left[-\frac{r^2}{2R_0^2(1 + iz/l_d)}\right]. \tag{34}$$

Comparing (34) with (28) and dropping the first term in the exponent in the latter expression by virtue of the inequality $\delta \ll 1$, we find ρ and Φ in trial solution (23):

$$\rho(r, z) = \rho_0\frac{R_0^2}{R^2}\exp\left(-\frac{r^2}{R^2}\right), \tag{35}$$

$$\Phi = -\sqrt{2b}\arctan\left(\frac{z}{l_d}\right) + \frac{zr^2}{2(l_d^2 + z^2)}, \tag{36}$$

where R is given by (33) with l replaced by l_d.

An analysis of expressions (23) and (36) shows that the paraxial portions of the electromagnetic pulse component move faster than its peripheral portions; i.e., the pulse is defocusing.

According to (33), the acousto-optic soliton exhibits diffractive spread analogous to that characteristic of quasi-monochromatic beams propagating in free space. Moreover, expressions (33)–(36), which describe the diffractive spread of the soliton, are similar to analogous expressions for quasi-monochromatic beams [24]. The pulse radius increases by a factor of $\sqrt{2}$ over the

distance l_d, while the amplitudes of its electromagnetic and acoustic components decrease by a factor of 2 and 4 according to (35) and (23), respectively. For the parameters of the pulse and optical material used above,

$$l_d \sim \omega R_0^2/c \sim 1 \text{ cm}.$$

The dynamics of a soliton with $2R_0 \sim 0.1$ cm is dominated by nonlinear refraction. Retaining only the second term in (29), we obtain $l = l_{def}$, where

$$l_{def} = R_0 \omega \tau_p \sqrt{n_m a/c}$$

is the defocusing length under nonlinear-refraction conditions. For the parameter values used above, $l_{def} \sim 10^3$ cm. This implies that the defocusing is sufficiently slow; i.e., soliton generation via SBSS can be observed experimentally.

5. CONCLUSIONS

The analysis presented above shows that forward Brillouin scattering of ultraslow light propagating at the speed of sound under conditions of electromagnetically induced transparency makes it possible to observe a new phenomenon: soliton generation via stimulated Brillouin self-scattering. The Stokes shift of an optical pulse can have any value within an interval, depending on the input conditions, rather than a certain value determined by the sound-to-light speed ratio and the input carrier frequency, as in the case of "classical" SBS. The shift increases with input pulse intensity, being limited from above by the condition that the optical-pulse spectrum remains within the EIT window.

This paper presents the simplest model of SBSS. As in the analyses of SRSS presented in [1, 4], the group-velocity dispersion parameter is assumed to be constant within the carrier bandwidth of the optical pulse. This assumption is justified by the fact that the center frequency of the EIT window corresponds to an inflection point in the frequency dependence of the resonant-impurity refractive index; i.e., its frequency derivative varies very slowly within the transparency window. In the general case, a numerical analysis of system (11), (12) should be performed, with $\chi(\tau)$ and $g(\tau)$ defined, respectively, as the Fourier preimages of the frequency-dependent susceptibility $\chi(\omega)$ and electrostriction coefficient $\gamma_2(\omega)$ within the EIT window.

SBSS is analogous to SRSS in that it does not involve any threshold condition. It is also important that SBSS can be observed when the input pulse intensity is as low as 10 W/cm^2, which is generally sufficient for manifestation of nonlinear effects under EIT conditions [7, 15, 16]. One fundamental distinction between SRSS and SBSS lies in the fact that the latter phenomenon can manifest itself only by soliton generation; i.e., dispersion is essential. Otherwise, forward SBS is impossible, whereas it is a key prerequisite for SBSS.

The Stokes shift in the optical-pulse frequency is associated with the generation of an acoustic video pulse, which evolves into an acousto-optic soliton. This effect is of both fundamental and applied interest: soliton generation via SBSS can be considered as a mechanism of optical-to-acoustic pulse conversion. Conversion mechanisms of this kind were analyzed for collinear (under EIT conditions) and noncollinear propagation of acoustic pulses and light in [25] and [26], respectively.

SBSS offers a mechanism for generation of optical solitons with tunable carrier frequency depending on input intensity, as in SRSS. The Stokes shift due to SBSS is on the order of 10^{-3}% of the carrier frequency, which is much smaller than that due to SRSS. Furthermore, the substantially lower intensity of the solitons generated via SBSS and the generation of carrier-free acoustic pulses are distinctive features of the phenomenon predicted in this study that have no analogs in SRSS.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 05-02-16422a.

REFERENCES

1. G. Agrawal, *Nonlinear Fiber Optics* (Academic, San Diego, 1995; Mir, Moscow, 1996).
2. R. H. Pantell and H. E. Puthoff, *Fundamentals of Quantum Electronics* (Wiley, New York, 1969; Mir, Moscow, 1972).
3. E. M. Dianov, A. Ya. Karasik, P. V. Mamyshev, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **41**, 242 (1985) [JETP Lett. **41**, 294 (1985)].
4. V. N. Serkin, T. L. Belyaeva, G. H. Corro, and M. Aguero Granados, Kvantovaya Élektron. (Moscow) **33**, 325 (2003).
5. L. V. Hau, S. E. Harris, Z. Dutton, and C. Behroozi, Nature **397**, 594 (1999).
6. A. V. Turukhin, V. S. Sudarshanam, M. S. Shahriar, *et al.*, Phys. Rev. Lett. **88**, 023602-1 (2002).
7. A. B. Matsko, Yu. V. Rostovtsev, M. Fleishhauer, and M. O. Scully, Phys. Rev. Lett. **86**, 2006 (2001).
8. S. E. Harris, Phys. Today, No. 6, 36 (1997).
9. A. Yariv, *Quantum Electronics*, 3rd ed. (Wiley, New York, 1989; Mir, Moscow, 1980).
10. S. A. Akhmanov, V. A. Vysloukh, and A. S. Chirkin, *The Optics of Femtosecond Laser Pulses* (Nauka, Moscow, 1988) [in Russian].
11. S. E. Harris, J. E. Field, and A. Kasapi, Phys. Rev. A **46**, R29 (1992).
12. V. E. Zakharov, Zh. Éksp. Teor. Fiz. **62**, 1745 (1972) [Sov. Phys. JETP **35**, 908 (1972)].

13. R. K. Dodd, J. C. Eilbeck, J. D. Gibbon, and H. C. Morris, *Solitons and Nonlinear Wave Equations* (Academic, London, 1984; Mir, Moscow, 1988).
14. N. Yadjima and M. Oikawa, Prog. Theor. Phys. **56**, 1719 (1976).
15. H. Schmidt and A. Imamoglu, Opt. Lett. **21**, 1936 (1996).
16. M. Lukin and A. Imamoglu, Phys. Rev. Lett. **84**, 1419 (2000).
17. S. K. Zhdanov and B. A. Trubnikov, Zh. Éksp. Teor. Fiz. **92**, 1612 (1987) [Sov. Phys. JETP **65**, 904 (1987)].
18. S. V. Sazonov, Zh. Éksp. Teor. Fiz. **125**, 1409 (2004) [JETP **98**, 1237 (2004)].
19. N. V. Karlov and N. A. Kirichenko, *Oscillations, Waves, and Structures* (Fizmatlit, Moscow, 2001) [in Russian].
20. I. I. Ol'khovskiĭ, *A Course of Theoretical Mechanics for Physicists* (Mosk. Gos. Univ., Moscow, 1978) [in Russian].
21. S. V. Sazonov, Zh. Éksp. Teor. Fiz. **119**, 419 (2001) [JETP **92**, 361 (2001)].
22. S. V. Sazonov, Usp. Fiz. Nauk **171**, 663 (2001) [Phys. Usp. **44**, 631 (2001)].
23. S. N. Vlasov, V. I. Talanov, and V. A. Petrishchev, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **14**, 1353 (1971).
24. S. A. Akhmanov and S. Yu. Nikitin, *Physical Optics*, 2nd ed. (Nauka/Mosk. Gos. Univ., Moscow, 2004; Clarendon, Oxford, 1997).
25. A. V. Gulakov and S. V. Sazonov, Pis'ma Zh. Éksp. Teor. Fiz. **79**, 746 (2004) [JETP Lett. **79**, 610 (2004)].
26. A. A. Zabolotskiĭ, Zh. Éksp. Teor. Fiz. **126**, 155 (2004) [JETP **99**, 133 (2004)].

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 101, No. 6, 2005, pp. 989–998.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 6, 2005, pp. 1134–1144.

ATOMS, MOLECULES, OPTICS

Probe Field Spectroscopy in Three-Level Λ Systems under Arbitrary Collisional Relaxation of Low-Frequency Coherence

A. I. Parkhomenko and A. M. Shalagin

Institute of Automation and Electrometry, Siberian Branch, Russian Academy of Sciences, Novosibirsk, 630090 Russia

e-mail: par@iae.nsk.su; shalagin@iae.nsk.su

Received April 26, 2005

Abstract—We investigate theoretically the spectrum of weak probe field absorption by three-level atoms with the Λ configuration of levels in the field of a strong electromagnetic wave acting on an adjacent transition and colliding with buffer gas atoms. Analysis is carried out for the general case of arbitrary collisional relaxation of low-frequency coherence at a transition between two lower levels. It is shown that, in the absence of collisional relaxation of low-frequency coherence, the probe field spectrum always exhibits clearly manifested anisotropy with respect to mutual orientation of wavevectors of the strong and probe radiation (even under small Doppler broadening). It is found that the probe field spectrum may acquire under certain conditions supernarrow resonances with a width proportional to the diffusion coefficient for atoms interacting with radiation. This fact may form the basis for a spectroscopic method for measuring transport frequencies of collisions between absorbing and buffer particles. A large-amplitude supernarrow resonance (with an amplitude much larger than the amplitude of the resonance near the line center), which is observed in the far wing of the absorption line, exhibits collisional narrowing (a nonlinear spectroscopic analog of the Dicke effect) at collision frequencies several orders of magnitude lower that the Doppler linewidth. Simple working equations proposed for describing the probe field spectrum are convenient for experimental data processing. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The probe field method in nonlinear spectroscopy is an effective tool for studying spectroscopic characteristics and various relaxation processes in quantum systems. The essence of the method is that a weak (probe) fields "probes" the structure of atomic states perturbed by another strong field [1–3].

A large number of publications are devoted to probe field spectroscopy. One of the most actively investigated systems is the three-level Λ system, in which transitions between the upper and each of the two lower levels of the system are dipole-allowed (see, for example, [1–10] and the literature cited therein). In their studies, most authors used simple relaxation models. In particular, it was assumed that complete phase mismatch is observed in induced polarization at all transitions during collisions leading to a change in the velocity. On the other hand, this assumption is not valid for a number of real experimental objects. For example, collisions of alkali metal atoms with nonmagnetic buffer particles do not break coherence between the hyperfine structure components of the ground electron state [11]. It was found that this circumstance leads to a radical change in the shape of spectral lines, which is undoubtedly important for solving a number of problems that have aroused a revival of interest (e.g., coherent capture of occupancies [12, 13], electromagnetically induced transparency [14], and lasing without population inversion [12–17]).

Here, we analyze theoretically the spectrum of probe field absorption by three-level atoms with the Λ configuration of the levels in the field of a strong electromagnetic wave acting on the adjacent transition. We assume that atoms are in the atmosphere of a buffer gas and experience collisions with its particles. Analysis is carried out for the general case of arbitrary collisional relaxation of low-frequency coherence at the transition between two lower levels (collisional relaxation may either be absent or, on the contrary, be quite effective). At the same time, we assume that collisions completely disturb the phase of the dipole moment induced by radiation at optical transitions between the common upper level and the lower levels.

Unexpectedly, the probe field spectrum turned out to be highly sensitive to collisions and motion of atoms even in cases when such factors can apparently be discarded. For example, the Dicke collisional narrowing effect for the probe field resonance in the far wing of the absorption line, which was discovered in the present study, may be strongly manifested at collision frequen-

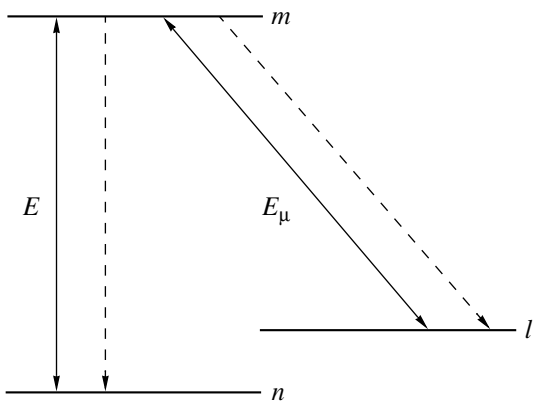


Fig. 1. Energy level diagram. Solid arrows denote transitions induced by radiation, while dashed arrows show spontaneous radiative transitions.

cies several orders of magnitude smaller than the Doppler width (it should be recalled that the Dicke effect in traditional spectroscopic problems [1, 2] is noticeably manifested only at collision frequencies larger than the Doppler width).

In the case of small Doppler broadening (relative to the collision frequency), the probe field spectrum should apparently be insensitive to the mutual orientation of the wavevectors of strong and probe radiations. Nevertheless, the probe field spectrum is strongly anisotropic to mutual orientation of wavevectors even in this case.

This study is devoted to analysis of these and other previously unknown features of the probe field spectrum of three-level Λ systems.

2. GENERAL EXPRESSIONS

Let us consider the interaction of strong and probe radiation with the gas of three-level absorbing particles mixed with a buffer gas. The energy level diagram of absorbing particles is shown in Fig. 1. Let a strong field

$$\mathrm{Re}\mathbf{E}\exp(i\mathbf{k}\cdot\mathbf{r}-i\omega t)$$

be in resonance with the *m–n* transition and a weak field

$$\mathrm{Re}\mathbf{E}_\mu\exp(i\mathbf{k}_\mu\cdot\mathbf{r}-i\omega_\mu t)$$

be in resonance with the adjacent *m–l* transition. The polarization of the medium at the probe field frequency is determined by the density matrix element $\rho_{ml}(\mathbf{v})$, where $\mathbf{v}$ is the particle velocity. These matrix elements can easily be found from the system of kinetic equations for density matrix elements (resonance approximation, see [1, 2])

$$\begin{aligned}&\left[\frac{d}{dt}+\frac{\Gamma_m}{2}-i(\Omega_{0\mu}-\mathbf{k}_\mu\cdot\mathbf{v})\right]\rho_{ml}(\mathbf{v})\\ &=S[\rho_{ml}(\mathbf{v})]+iG_\mu[\rho_{ll}(\mathbf{v})-\rho_{mm}(\mathbf{v})]+iG\rho_{nl}(\mathbf{v}),\\ &\left\{\frac{d}{dt}+i[\Omega_0-\Omega_{0\mu}+(\mathbf{k}_\mu-\mathbf{k})\cdot\mathbf{v}]\right\}\rho_{nl}(\mathbf{v})\\ &=S[\rho_{nl}(\mathbf{v})]+iG^*\rho_{ml}(\mathbf{v})-iG_\mu\rho_{nm}(\mathbf{v}),\end{aligned}\qquad(1)$$

where

$$\begin{aligned}\Omega_0&=\omega-\omega_{mn},\quad \Omega_{0\mu}=\omega_\mu-\omega_{ml},\\ G&=\frac{d_{mn}E}{2\hbar},\quad G_\mu=\frac{d_{ml}E_\mu}{2\hbar}.\end{aligned}\qquad(2)$$

Here, $\rho_{ii}(\mathbf{v})$ is the velocity distribution of particles at the ith level ($i = m, n$); $S[\rho_{ij}(\mathbf{v})]$ are collision integrals; Γ_m is the total spontaneous rate of decay of the excited level m (via the $m \longrightarrow n$ and $m \longrightarrow l$ channels); d_{mn} and d_{ml} are the matrix elements of the dipole moments of the *m–n* and *m–l* transitions; and ω_{mn} and ω_{ml} are the frequencies of the *m–n* and *m–l* transitions.

We will consider the case when collisions completely disturb the phase of the dipole moment induced at the *m–l* transition, but when collisional relaxation of low-frequency coherence $\rho_{nl}(\mathbf{v})$ is arbitrary. We assume that collision integral $S[\rho_{ml}(\mathbf{v})]$ in relation (1) satisfies the conventional approximation for the present case [1, 3]

$$S[\rho_{ml}(\mathbf{v})]=-\nu_{ml}\rho_{ml}(\mathbf{v}),\qquad(3)$$

where the "departure" frequency ν_{ml} is a complex quantity in the general case. For the collision integral $S[\rho_{nl}(\mathbf{v})]$, we will use the model of strong collisions [1],

$$\begin{aligned}S[\rho_{nl}(\mathbf{v})]&=-\nu\rho_{nl}(\mathbf{v})+\tilde{\nu}\rho_{nl}W(\mathbf{v}),\\ \rho_{nl}&\equiv\int\rho_{nl}(\mathbf{v})d\mathbf{v},\end{aligned}\qquad(4)$$

where $W(\mathbf{v})$ is the Maxwell velocity distribution and ν and $\tilde{\nu}$ are the "departure" and "arrival" frequencies, which are generally complex quantities. The value $\tilde{\nu} = 0$ corresponds to the absence of phase memory in collisions (collisions induce complete relaxation of coherence $\rho_{nl}(\mathbf{v})$). In the case of absolute phase memory in collisions (the absence of collisional relaxation of coherence $\rho_{nl}(\mathbf{v})$), the frequencies ν and $\tilde{\nu}$ of departure and arrival are real-valued and identical [1],

$$\tilde{\nu}=\nu=\nu_{\mathrm{tr}},\qquad(5)$$

where ν_{tr} has the meaning of the average transport frequency of elastic collisions between active and buffer

particles [18]. Quantity ν_{tr} is connected with diffusion coefficient D for particles interacting with radiation via the relation [19]

$$D = v_T^2/2\nu_{tr}, \quad (6)$$

where v_T is the most probable velocity of absorbing particles.

Under steady-state and spatially homogeneous conditions, we obtain the following relations from Eq. (1) combined with (3) and (4):

$$\begin{aligned}\lambda_1(\mathbf{v})\rho_{ml}(\mathbf{v}) &= iG_\mu[\rho_{ll}(\mathbf{v}) - \rho_{mm}(\mathbf{v})] + iG\rho_{nl}(\mathbf{v}),\\ \lambda_2(\mathbf{v})\rho_{nl}(\mathbf{v}) &= \tilde{\nu}\rho_{nl}W(\mathbf{v}) - iG_\mu\rho_{nm}(\mathbf{v})\\ &\quad + iG^*\rho_{ml}(\mathbf{v}),\end{aligned} \quad (7)$$

where

$$\begin{aligned}\lambda_1(\mathbf{v}) &= \frac{\Gamma_m}{2} + \nu_{ml} - i(\Omega_{0\mu} - \mathbf{k}_\mu\cdot\mathbf{v}),\\ \lambda_2(\mathbf{v}) &= \nu + i(\Omega_0 - \Omega_{0\mu} + \mathbf{q}\cdot\mathbf{v}),\\ \mathbf{q} &\equiv \mathbf{k}_\mu - \mathbf{k}, \quad \rho_{ij} \equiv \int\rho_{ij}(\mathbf{v})d\mathbf{v}.\end{aligned} \quad (8)$$

In view of the low intensity of the probe field, matrix elements $\rho_{ii}(\mathbf{v})$ and $\rho_{nm}(\mathbf{v})$ in Eqs. (7) can be assumed to be known and determined by the action of the strong field alone. Further, we assume that collisional transitions between levels n and l are absent. In this case, all particles pass from level n to level l under the action of the strong field. Consequently, in Eqs. (7) we can set

$$\rho_{mm}(\mathbf{v}) = \rho_{nm}(\mathbf{v}) = 0, \quad \rho_{ll}(\mathbf{v}) = NW(\mathbf{v}), \quad (9)$$

where N is the total concentration of absorbing particles.

Here, we will analyze the probe field absorption spectrum. Using system of equations (7) combined with relations (9), we derive the following expression for the probability P_μ of probe field absorption at frequency ω_μ (the number of radiation absorption events per unit time for an absorbing atom):

$$\begin{aligned}P_\mu &\equiv -\frac{2}{N}\mathrm{Re}[iG_\mu^*\rho_{ml}]\\ &= 2|G_\mu|^2\mathrm{Re}\left\{I_2 - \frac{\tilde{\nu}|G|^2J^2}{1-\tilde{\nu}I_1}\right\},\end{aligned} \quad (10)$$

where

$$\begin{aligned}J &= \int\frac{W(\mathbf{v})d\mathbf{v}}{\lambda_1(\mathbf{v})\lambda_2(\mathbf{v}) + |G|^2},\\ I_i &= \int\frac{\lambda_i(\mathbf{v})W(\mathbf{v})d\mathbf{v}}{\lambda_1(\mathbf{v})\lambda_2(\mathbf{v}) + |G|^2}, \quad i = 1, 2.\end{aligned} \quad (11)$$

Thus, the calculation of the probe field spectrum in the model of strong collisions is reduced to evaluation of the corresponding integrals.

3. ANALYSIS OF THE PROBE FIELD SPECTRUM

It is difficult to analyze expression (10) in the general form, although the specific probe field spectrum for various sets of parameters (intensity of strong radiation, its frequency, collision frequencies, and the extent of manifestation of the phase memory) can easily be obtained from numerical calculations. To clarify the physical pattern, we will first consider various special cases for which relatively simple analytic results can be obtained from expression (10).

3.1. Homogeneous Broadening

We begin our analysis from the case of homogeneous broadening of the absorption line at the *m–l* transition, when the Doppler width $k_\mu v_T$ is smaller than Γ:

$$\Gamma \gg k_\mu v_T, \quad \Gamma = \frac{\Gamma_m}{2} + \mathrm{Re}\nu_{ml}. \quad (12)$$

In this case, for not very high radiation intensity, such that

$$|G|^2 \ll |\nu(\Gamma - i\Omega)|, \quad (13)$$

formula (10) for the probe field absorption probability can be substantially simplified and assumes the form[1]

$$P_\mu = 2|G_\mu|^2\mathrm{Re}\left\{\left(\Gamma - i\Omega_\mu + \frac{|G|^2}{\Gamma_1 - i\varepsilon}\right)^{-1}\right\}. \quad (14)$$

Here, the following notation has been introduced:

$$\begin{aligned}\Gamma_1 &= (\nu - \tilde{\nu})' + \frac{(qv_T)^2}{2[\nu + i(\Omega_0 - \Omega_{0\mu})]}, \quad \varepsilon = \Omega_\mu - \Omega,\\ \Omega_\mu &= \Omega_{0\mu} - \nu_{ml}'', \quad \Omega = \Omega_0 + (\nu - \tilde{\nu} - \nu_{ml})'',\\ q &\equiv |\mathbf{k}_\mu - \mathbf{k}|.\end{aligned} \quad (15)$$

Here and below, prime and double prime indicate the real and imaginary parts of a complex number, respectively.[2]

Formula (14) for the absorption probability is similar in structure to the corresponding formula for particles at rest (see, for example, [1, 2]). The only differ-

[1] It should be noted that condition (13) depends on radiation frequency detuning Ω and can be satisfied even for $|G| \gtrsim \Gamma$ for large detunings ($|\Omega| \gg \Gamma$).

[2] A formula analogous to expression (14) was recently derived in [10], where the probe field spectrum for a three-level Λ system is considered under possible conservation of phase memory during collisions at all atomic transitions simultaneously. The main attention in [10] was paid to analysis of the shape and positions of the components of the Autler-Townes doublet associated with field splitting of the upper level in the Λ system.

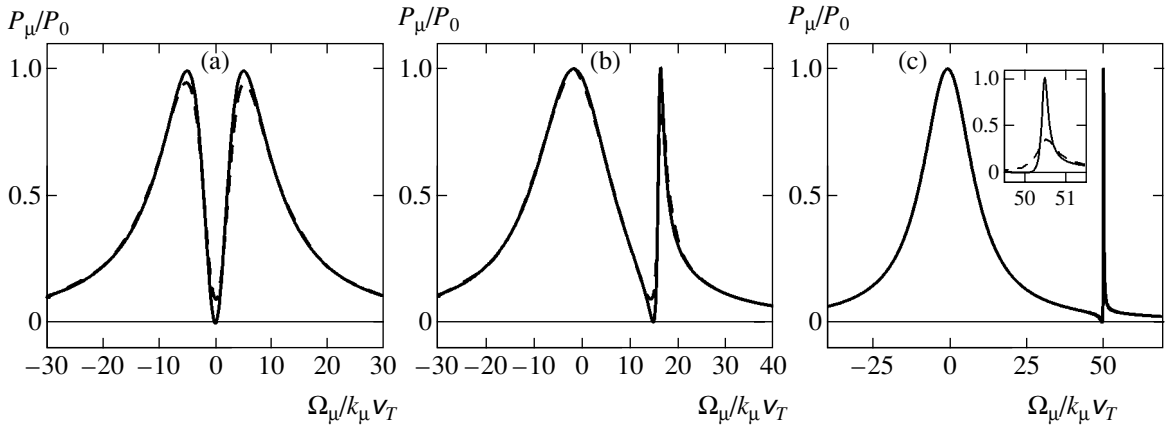


Fig. 2. Dependences of the probe field absorption probability P_μ on the frequency detuning Ω_μ in the case of homogeneous broadening of the absorption line at the *m–l* transition ($\Gamma \gg k_\mu v_T$) and in the absence of collisional relaxation of low-frequency coherence at the *n–l* transition ($\tilde{\nu} = \nu$), $|G|/kv_T = 5$, $\nu'/k_\mu v_T = 10$; $\Gamma_m/k_\mu v_T = 10^{-2}$, and $(k - k_\mu)/k = 10^{-4}$; solid and dashed curves correspond to $\mathbf{k}_\mu \uparrow\uparrow \mathbf{k}$ and $\mathbf{k}_\mu \uparrow\downarrow \mathbf{k}$, respectively; $\Omega = 0$ (a), $\Omega/kv_T = 15$ (b), and 50 (c); the inset shows the resonance in the vicinity of $\Omega_\mu = \Omega_\mu^{(+)} \approx \Omega + |G|^2/\Omega$ on a magnified scale.

ence is that the relaxation constant Γ_1 should be generalized to take into account the diffusion motion of particles (second term in formula (15) for Γ_1). In the case of phase memory conservation in collisions at the *n–l* transitions, relaxation constant Γ_1 is determined only by the second (diffusion) term. Thus, motion of atoms should always be taken into account in analysis of the probe field spectrum, even in the case of small Doppler broadening relative to the collision frequency.

The quantity Γ_1 is anisotropic to mutual orientation of wavevectors of strong and probe radiation. If phase memory is preserved at the n–*l* transition and the magnitudes of the wavevectors differ insignificantly ($|k - k_\mu| \ll k$), the value of Γ_1 for counterpropagating ($\mathbf{k}_\mu \uparrow\downarrow \mathbf{k}$) and unidirectional ($\mathbf{k}_\mu \uparrow\uparrow \mathbf{k}$) waves may differ by many orders of magnitude. This means that the probe field spectrum exhibits a clearly manifested anisotropy to mutual orientation of the wavevectors of the strong and probe radiations even in the limit of small Doppler broadening.

Considering the limiting case $\Gamma_1 \longrightarrow 0$ in formula (14), we can easily note that the maximal values of P_μ are attained under the condition

$$\varepsilon\Omega_\mu = |G|^2.$$

This condition leads to the well-known result (see, for example, [1, 2]) that the probe field spectrum has two components whose peaks lie in the vicinity of $\Omega_\mu = \Omega_\mu^{(\pm)}$, where

$$\Omega_\mu^{(\pm)} = \frac{1}{2}(\Omega \pm \Omega_R), \quad \Omega_R \equiv \sqrt{4|G|^2 + \Omega^2}. \qquad (16)$$

In accordance with these formulas, the distance between the spectral component peaks is Ω_R. In spite of the limitations used in deriving formula (16), it defines the position of spectral components to a high degree of accuracy for any values of the parameters of the problem.

Figure 2 shows the general form of the probe field spectrum for various detunings Ω of the strong field frequency in the case of homogeneous broadening (12) and phase memory conservation at the *n–l* transition. In the case of exact resonance for the strong field ($\Omega = 0$), a "negative" symmetric structure exists at the line center (dip in Fig. 2a). With increasing $|\Omega|$, a narrow resonance peak appears in the wing of the absorption line in the vicinity of $\Omega_\mu = \Omega_\mu^{(+)}$ (Figs. 2b and 2c). The behavior of this resonance is quite peculiar: its amplitude remains unchanged and equal to the amplitude of the resonance near the absorption line center even for a large detuning $|\Omega_\mu| \gg \Gamma$ (far wing of the absorption line; see Fig. 2c); the resonance width rapidly decreases with increasing $|\Omega|$ and may be smaller than the natural linewidth Γ_m. In the vicinity of $\Omega_\mu = \Omega$, the probe field spectrum acquires a dip (see Fig. 2); absorption of radiation becomes much weaker and practically vanishes under certain conditions ("dark resonance" [12, 13]). In all figures, the unit of measurement is the quantity P_0, viz., the absorption probability for probe radiation at the line center at the *m–l* transition in the limit of a low intensity of the strong field. In accordance with formula (10), we have

$$P_0 = 2|G_\mu|^2 \mathrm{Re}\int \frac{W(\mathbf{v})d\mathbf{v}}{\lambda_1(\mathbf{v})}. \qquad (17)$$

In the case of homogeneous broadening of the absorption line ($\Gamma \gg k_\mu v_T$), we have

$$P_0 = \frac{2|G_\mu|^2}{\Gamma},$$

in the case of Doppler broadening ($\Gamma \ll k_\mu v_T$), we have

$$P_0 = \frac{2\sqrt{\pi}|G_\mu|^2}{k_\mu v_T}.$$

Let us consider in greater detail various regions of the probe field spectrum. For a small strong field frequency detuning

$$|\Omega| \ll \Gamma, \tag{18}$$

it follows from formula (14) that the probe field spectrum can be described by the formula[3]

$$P_\mu = 2|G_\mu|^2 \times \mathrm{Re}\left\{\frac{1}{\Gamma - i\Omega_\mu} - \frac{|G|^2}{\Gamma^2\left[\Gamma_1 + \frac{|G|^2}{\Gamma} - i(\Omega_\mu - \Omega)\right]}\right\}. \tag{19}$$

In accordance with this formula, the probe field spectrum contains two Lorentz profiles, which are summed with different signs. The first term in the braces describes the Lorentz absorption line of width Γ, which is typical of the case with an extremely low intensity of the strong field. The second term describes the dip in the vicinity of $\Omega_\mu = \Omega$ with the half-width $\Gamma_1 + |G|^2/\Gamma$ against the background of the Lorentz profile with half-width Γ, which is associated with the first term. The dip width is anisotropic to mutual orientation of the wavevectors of the strong and probe radiations (Fig. 3a). It can be seen from the figure that the resonance in the case of counterpropagating waves is broadened to such an extent that the dashed curve is almost horizontal.

Anisotropy is manifested most strongly in the case of a low radiation intensity ($|G|^2 \ll \Gamma\Gamma_1$), when the half-width of the dip is equal to Γ_1. In the case of phase memory conservation at the *n*–*l* transition, quantity Γ_1 in formula (19) satisfies the following expression:

$$\Gamma_1 = q^2 D. \tag{20}$$

[3] In the formulas for the probe field spectrum, quantity Γ_1 (15) can be assumed to be real-valued (we can carry out the substitution $\Gamma_1 \longrightarrow \mathrm{Re}\Gamma_1$) under the condition $\nu' \gg qv_T$, which is in fact always satisfied in the case (12) of homogeneous broadening of the absorption line).

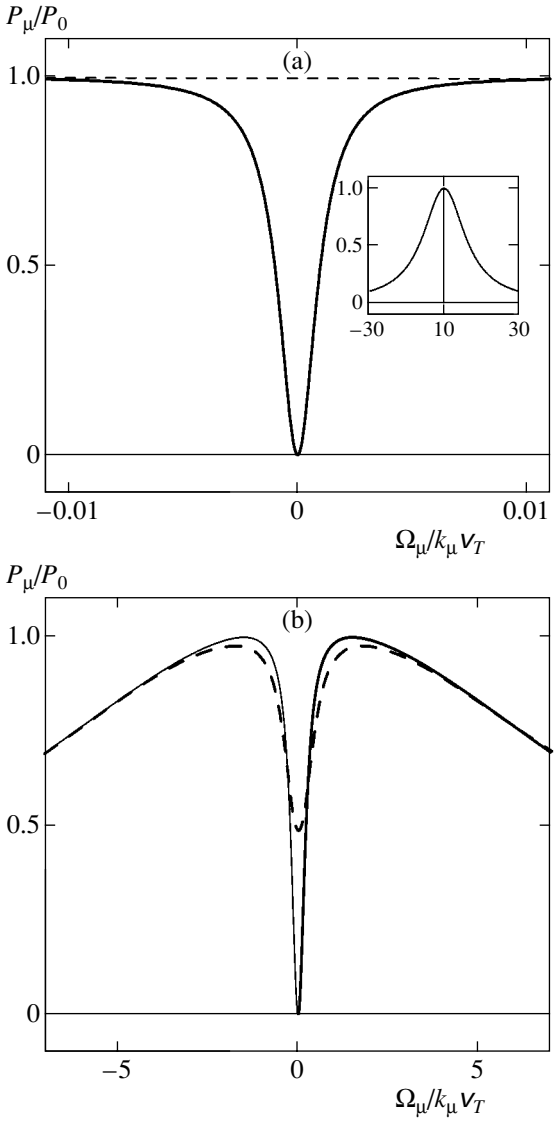


Fig. 3. Dependences $P_\mu(\Omega_\mu)$ in the case of homogeneous broadening for exact resonance for the strong field ($\Omega = 0$); collisions preserve phase memory at the *n*–*l* transition, $\nu'/k_\mu v_T = 10$, $(k - k_\mu)/k = 10^{-4}$, $\Gamma_m/k_\mu v_T = 10^{-2}$; solid and dashed curves correspond to $\mathbf{k}_\mu \uparrow\uparrow \mathbf{k}$ and $\mathbf{k}_\mu \uparrow\downarrow \mathbf{k}$, respectively; $|G|/kv_T = 0.1$ (a; the inset shows the general view of the spectrum) and 1.5 (b).

Thus, the dip width in this case turns out to be proportional to diffusion coefficient *D* for particles interacting with radiation. This circumstance may serve as the basis for a spectroscopic method for measuring the diffusion coefficient of absorbing particles in the buffer gas atmosphere.

With increasing radiation intensity (for $|G|^2 \sim \Gamma\Gamma_1$), the widths of the dips for unidirectional and counterpropagating waves become comparable; however, in the latter case no appreciable bleaching of the medium takes place (Fig. 3b). As the radiation intensity

increases further (for $|G|^2 \gg \Gamma\Gamma_1$), the anisotropy of the probe field spectrum practically disappears (see Fig. 2a).

Let us consider the case of extremely strong detuning of the strong field frequency:

$$|\Omega| \gg \Gamma. \tag{21}$$

In this case, formula (14) for the probe field absorption probability can be transformed to

$$P_\mu = \frac{2|G_\mu|^2}{\Gamma}\left\{\frac{\Gamma^2}{\Gamma^2 + (\Omega_\mu + |G|^2/\Omega)^2} + \frac{(\Gamma_1 + \gamma)\gamma}{(\Gamma_1 + \gamma)^2 + (\Omega_\mu - \Omega - |G|^2/\Omega)^2}\right\}, \tag{22}$$

$$\gamma = \frac{|G|^2\Gamma}{\Omega^2}.$$

In accordance with this formula, the probe field spectrum consists of two Lorentz profiles with half-widths Γ and $\Gamma_1 + \gamma$, which are located in the vicinity of

$$\Omega_\mu = \Omega_\mu^{(-)} \approx -\frac{|G|^2}{\Omega}$$

and

$$\Omega_\mu = \Omega_\mu^{(+)} \approx \Omega + \frac{|G|^2}{\Omega},$$

respectively.

The width $2(\Gamma_1 + \gamma)$ of the resonance in the far wing of the line (in the vicinity of $\Omega_\mu \approx \Omega$) is anisotropic to mutual orientation of the wavevectors of the strong and probe radiations (see the inset to Fig. 2c) and may be smaller than the natural linewidth Γ_m.

If phase memory is preserved during the *n–l* transition, the quantity Γ_1 is defined by formula (20) and, hence, the resonance width depends on the diffusion coefficient D of particles interacting with radiation. The ratio a of the amplitude of the resonance in the far wing of the line to the amplitude of the resonance in the vicinity of the line center is given by

$$a = \frac{\gamma}{\Gamma_1 + \gamma} \tag{23}$$

and is close to unity for $\Gamma_1 \ll \gamma$ (see Fig. 2c). Thus, for $\Gamma_1 \ll \Gamma$, the amplitude of the resonance in the far wing of the line does not decrease with increasing strong field frequency detuning up to values of

$$|\Omega| \sim |G|\sqrt{\Gamma/\Gamma_1}.$$

In the absence of phase memory at the *n–l* transition, we have $\Gamma_1 \sim \Gamma \gg \gamma$ and, hence, the resonance in the wing of the line is weakly pronounced (its relative amplitude a is small).

Expressing quantity Γ_1 from Eq. (23), we arrive at the relation

$$\frac{\tilde{\nu}'}{\nu'} = 1 + \frac{(qv_T)^2}{2\nu'^2} - \frac{\gamma}{\nu'}\frac{1-a}{a} \approx 1 + \frac{2D}{v_T^2}\left[q^2 D - \frac{\gamma(1-a)}{a}\right], \tag{24}$$

which can be used to find the extent of phase memory conservation in collisions, characterized by parameter $\tilde{\nu}'/\nu'$ ($0 \le \tilde{\nu}'/\nu' \le 1$) from the relative amplitude a of the resonance in the line wing. The second approximate equality in formula (24) (with diffusion coefficient D on the right-hand side) is in fact exact in the case of a high extent of phase memory conservation at the *n–l* transition (for $1 - \tilde{\nu}'/\nu' \ll 1$) since we can set

$$\nu' = \nu_{tr} = v_T^2/2D$$

on the right-hand side of the first equality in this case.

3.2. Doppler Broadening

Let us now analyze the case of Doppler broadening of the absorption line at the *m–l* transition ($k_\mu v_T \gg \Gamma$). Figure 4 shows typical probe field spectra for various values of detuning Ω. As compared to the case of homogeneous broadening considered earlier, the line profile $P_\mu(\Omega_\mu)$ experiences considerable changes: the spectrum anisotropy increases, and the amplitude of the narrow resonance in the absorption line wing (in the vicinity of $\Omega_\mu = \Omega_\mu^{(+)}$ increases with detuning Ω (Figs. 4b and 4c) and is found to be much larger than the amplitude of the resonance near the line center when $|\Omega| \gg \Gamma$ (see Fig. 4c).

Let us consider in detail the most interesting case of Doppler broadening for a large detuning of the strong radiation frequency,

$$|\Omega| \gg k_\mu v_T \gg \Gamma. \tag{25}$$

Let us suppose that the radiation intensity is not too high (see relation (13)) and the following conditions are satisfied:

$$\nu' \gg qv_T, \quad k_\mu v_T \frac{|G|^2}{\Omega^2}, \quad \text{for} \quad \mathbf{q} + \mathbf{k}_\mu \frac{|G|^2}{\Omega^2} \neq 0, \tag{26}$$

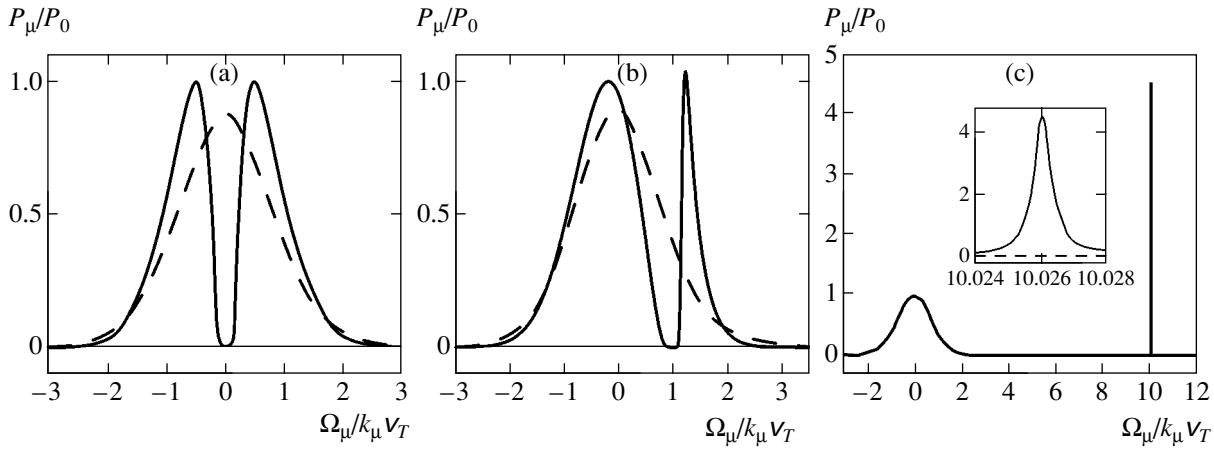


Fig. 4. Dependences $P_\mu(\Omega_\mu)$ in the case of Doppler broadening of the absorption line at the m–l transition ($k_\mu v_T \gg \Gamma$); collisions preserve phase memory at the n–l transition ($\tilde{\nu} = \nu$), $|G|/kv_T = 0.5$, $\nu'/k_\mu v_T = 10^{-2}$, $\Gamma_m/k_\mu v_T = 10^{-2}$; $(k - k_\mu)/k = 10^{-4}$; solid and dashed curves correspond to $\mathbf{k}_\mu \uparrow\uparrow \mathbf{k}$ and $\mathbf{k}_\mu \uparrow\downarrow \mathbf{k}$, respectively; $\Omega = 0$ (a), $\Omega/kv_T = 1$ (b), and 10 (c); the inset shows the resonance in the vicinity of $\Omega_\mu = \Omega_\mu^{(+)} \approx \Omega + |G|^2/\Omega$ on a magnified scale.

and

$$|\Omega|\Gamma \gg (k_\mu v_T)^2 \quad \text{or} \quad \nu' \gg (k_\mu v_T)^2 \frac{|G|^2}{|\Omega|^3}, \qquad \text{for} \quad \mathbf{q} + \mathbf{k}_\mu \frac{|G|^2}{\Omega^2} = 0. \tag{27}$$

We can prove that formula (22) for the absorption probability under conditions (25)–(27) correctly describes the resonance in the line wing (in the vicinity of $\Omega_\mu = \Omega + |G|^2/\Omega$, the second term in formula (22)). However, the resonance at the line center (in the vicinity of $\Omega_\mu = -|G|^2/\Omega$) cannot be described by formula (22); this resonance should be analyzed proceeding from the general formula (10).

To analyze the resonance in the vicinity of the absorption line center, let us consider the special case of equality of the wavevectors of the strong and probing radiation ($\mathbf{k}_\mu = \mathbf{k}$). In this case, formula (10) for P_μ can be substantially simplified and assumes the form

$$P_\mu = \frac{2\sqrt{\pi}|G_\mu|^2}{k_\mu v_T} \times \mathrm{Re}\left\{\frac{w(b)}{1 - \dfrac{\sqrt{\pi}\tilde{\nu}|G|^2 w(b)}{k_\mu v_T[\varepsilon + i(\nu - \tilde{\nu})'](\varepsilon_1 + i\nu')}}\right\}, \tag{28}$$

where

$$w(b) \equiv \frac{i}{\pi}\int_{-\infty}^{\infty} \frac{\exp(-t^2)}{b - t}dt,$$

$$b = \frac{1}{k_\mu v_T}\left[\left(\Omega_\mu - \frac{|G^2|\varepsilon_1}{\nu'^2 + \varepsilon_1^2}\right) + i\left(\Gamma + \frac{|G|^2\nu'}{\nu'^2 + \varepsilon_1^2}\right)\right], \tag{29}$$

$$\varepsilon_1 = \varepsilon - \tilde{\nu}'',$$

$w(b)$ being the probability integral of a complex argument, which is tabulated in [20].

Under Doppler broadening conditions for the absorption line ($\Gamma \ll k_\mu v_T$), parameter b in formula (28) may be small ($|b| \lesssim 1$). In this case, we can set

$$w(b) \approx \exp(-b^2),$$

which immediately leads to the following expression following from formula (28) in the absence of collisions ($\nu, \tilde{\nu} = 0$):

$$P_\mu = \frac{2\sqrt{\pi}|G_\mu|^2}{k_\mu v_T} \times \exp\left\{-\left[\Omega_\mu - \frac{|G|^2}{\Omega_\mu - \Omega}\right]^2 \frac{1}{(k_\mu v_T)^2}\right\}. \tag{30}$$

Numerical analysis shows that, in the absence of collisions, formula (30) is a good approximation of the exact, but more cumbersome expression (28) for any value of parameter b. This is due to the fact that the con-

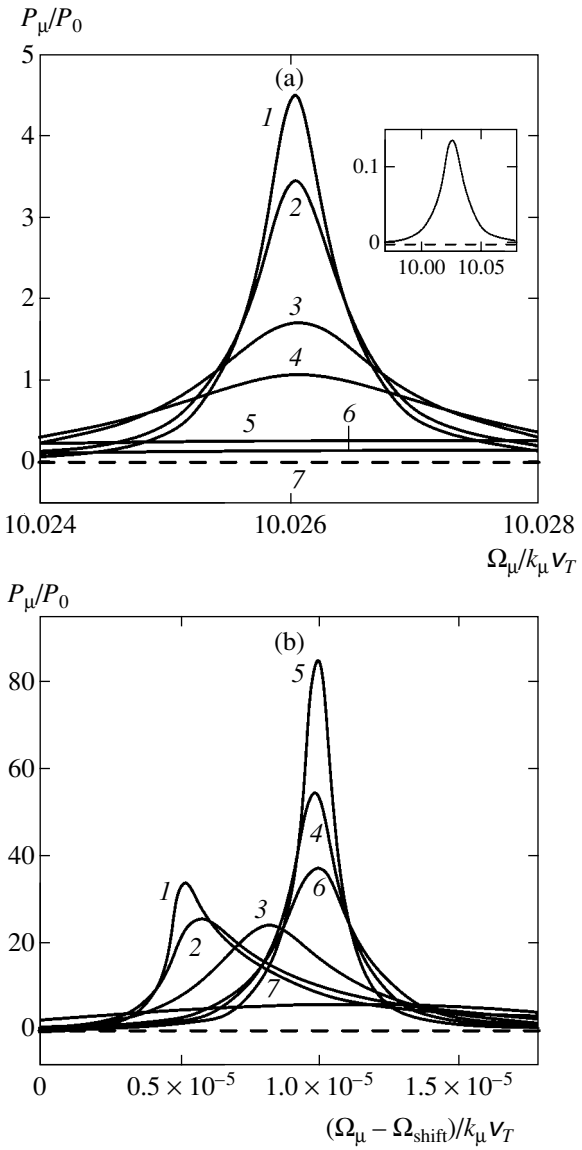


Fig. 5. Dependences $P_\mu(\Omega_\mu)$ in the vicinity of $\Omega_\mu = \Omega + |G|^2/\Omega$ in the case of Doppler broadening; $\Omega/kv_T = 10$, $\Gamma_m/k_\mu v_T = 10^{-2}$; $(k - k_\mu)/k = 10^{-4}$; solid and dashed curves correspond to $\mathbf{k}_\mu \uparrow\uparrow \mathbf{k}$ and $\mathbf{k}_\mu \uparrow\downarrow \mathbf{k}$, respectively; (a) $|G|/kv_T = 0.5$, $\nu'/k_\mu v_T = 10^{-2}$; $\tilde{\nu} = \nu$ (*1*), $\tilde{\nu}'/\nu' = 0.99$ (*2*), 0.95 (*3*), 0.9 (*4*), 0.5 (*5*); $\tilde{\nu} = 0$ (*6*); $\mathbf{k}_\mu \uparrow\downarrow \mathbf{k}$ for any ratio $\tilde{\nu}/\nu$ (*7*); the inset shows curves *6* and *7* on a magnified scale; (b) $|G|/kv_T = 0.1$ ($|G| = |\Omega|\sqrt{q/k_\mu}$), $\Omega_{shift}/k_\mu v_T = 10.002$; $\nu = 0$ (*1*); $\nu'/kv_T = 10^{-6}$ (*2*), 10^{-5} (*3*), 10^{-4} (*4*); 10^{-3} (*5*); 10^{-2} (*6*), 10^{-1} (*7*); $\mathbf{k}_\mu \uparrow\downarrow \mathbf{k}$ for any ratio ν'/kv_T (*8*).

dition $|b| \lesssim 1$ of applicability of this formula in the vicinity of spectral components (near $\Omega_\mu = \Omega_\mu^{(\pm)}$) is always satisfied (the exponent in formula (30) vanishes for $\Omega_\mu = \Omega_\mu^{(\pm)}$). However, in the detuning range, where $|b| \gg 1$, the absorption probability is low (as compared to the unit of measurement P_0) and, hence, the difference between the results of calculation by formulas (28) and (30) is small.

In addition, numerical analysis shows that for a large detuning of the strong radiation frequency (25), formula (30) correctly describes the resonance near the line center in the presence of collisions and for any relation between $\mathbf{k}_\mu$ and $\mathbf{k}$ as well, in spite of the fact that this formula was derived for the special case when $\mathbf{k}_\mu = \mathbf{k}$ and $\nu, \tilde{\nu} = 0$. Summarizing the above arguments, we conclude that the probe field absorption probability for a large value of $|\Omega|$ in the case of Doppler broadening (under conditions (25)–(27)) is described by the formula

$$P_\mu = \frac{2|G_\mu|^2}{\Gamma}\left\{\frac{\sqrt{\pi}\Gamma}{k_\mu v_T}\exp\left[-\left(\frac{\Omega_\mu + |G|^2/\Omega}{k_\mu v_T}\right)^2\right] + \frac{(\Gamma_1 + \gamma)\gamma}{(\Gamma_1 + \gamma)^2 + (\Omega_\mu - \Omega - |G|^2/\Omega)^2}\right\}. \quad (31)$$

In accordance with this formula, the probe field spectrum consists of a Lorentz profile with half-width $\Gamma_1 + \gamma$, which is located in the far wing of the line (in the vicinity of $\Omega_\mu = \Omega + |G|^2/\Omega$) and a Doppler profile of half-width $k_\mu v_T$, which is located near the line center (in the vicinity of $\Omega_\mu = -|G|^2/\Omega$). The ratio A of the amplitude of the resonance in the far wing of the line to the amplitude of the resonance near the line center is given by

$$A = \frac{k_\mu v_T}{\sqrt{\pi}\Gamma}\frac{\gamma}{\Gamma_1 + \gamma} \quad (32)$$

and can be much larger than unity (see Fig. 4c).

Expression (32) can be used for deriving a relation connecting the relative amplitude A of the resonance in the line wing with the degree of phase memory conservation in collisions:

$$\frac{\tilde{\nu}'}{\nu'} = 1 + \frac{(qv_T)^2}{2\nu'^2}\frac{1}{1 + |G|^4/\Omega^2\nu'^2} - \frac{\gamma}{\nu'}\left(\frac{k_\mu v_T}{\sqrt{\pi}\Gamma A} - 1\right). \quad (33)$$

For a high degree of phase memory conservation at the n–l transition (for $1 - \tilde{\nu}'/\nu' \ll 1$), the collision frequency ν' on the right-hand side of Eq. (33) can be expressed in terms of diffusion coefficient D:

$$\nu' = \nu_{tr} = v_T^2/2D.$$

Pay attention to the fact that the amplitude and width of the resonance in the line wing are strongly sensitive

to phase memory effects in collisions in spite of the smallness of the collision frequency as compared to the Doppler width (Fig. 5a).

It is well known (see, for example, [1, 2]) that the Dicke effect is noticeably manifested in traditional spectroscopic problems only when the collision frequency is much larger than the Doppler linewidth. However, in the case considered here, the Dicke effect is strongly manifested even for a collision frequency much smaller than the Doppler width. The physical reason for this effect can easily be grasped using the concepts of field-induced level splitting (such a splitting is also known as the Autler–Townes effect, or the dynamic Stark effect).

In a strong field, level m with unperturbed energy $E_m^{(0)}$ splits into two sublevels with quasi-energies $E_m^{(\pm)}$. For particles moving at a velocity $\mathbf{v}$, quasi-energy $E_m^{(\pm)}$ is a function of $\mathbf{v}$ due to the fact that the frequency of the field acting on a particle depends on its velocity:

$$\begin{aligned} E_m^{(\pm)}(\mathbf{v}) &= E_m^{(0)} + \hbar\Omega_\mu^{(\pm)}(\mathbf{v}), \\ \Omega_\mu^{(\pm)}(\mathbf{v}) &= \frac{1}{2}[\Omega - \mathbf{kv} \pm \sqrt{4|G|^2 + (\Omega - \mathbf{k}\cdot\mathbf{v})^2}]. \end{aligned} \tag{34}$$

For particles with a fixed velocity $\mathbf{v}$, detuning $\Delta_\mu^{(\pm)}(\mathbf{v})$ of the probe field frequency at the transition between the sublevel with quasi-energy $E_m^{(\pm)}(\mathbf{v})$ and level l with energy $E_l^{(0)}$ is given by

$$\begin{aligned} \Delta_\mu^{(\pm)}(\mathbf{v}) &\equiv \omega_\mu - \mathbf{k}_\mu\cdot\mathbf{v} - \frac{E_m^{(\pm)}(\mathbf{v}) - E_l^{(0)}}{\hbar} \\ &= \Omega_\mu - \mathbf{k}_\mu\cdot\mathbf{v} - \Omega_\mu^{(\pm)}(\mathbf{v}). \end{aligned} \tag{35}$$

For $|G| \ll |\Omega|$, we obtain

$$\begin{aligned} \Delta_\mu^{(+)}(\mathbf{v}) &= \Omega_\mu - \Omega - \frac{|G|^2}{\Omega} - \mathbf{q}\cdot\mathbf{v} - \mathbf{k}\cdot\mathbf{v}\frac{|G|^2}{\Omega^2}, \\ \Delta_\mu^{(-)}(\mathbf{v}) &= \Omega_\mu + \frac{|G|^2}{\Omega} - \mathbf{k}_\mu\cdot\mathbf{v} + \mathbf{k}\cdot\mathbf{v}\frac{|G|^2}{\Omega^2}. \end{aligned} \tag{36}$$

For an insignificant difference between the wavevectors ($\mathbf{k}_\mu \approx \mathbf{k}$, $q \ll k$), the Doppler shift in detuning $\Delta_\mu^{(+)}(\mathbf{v})$ is small as compared to $\mathbf{k}_\mu\cdot\mathbf{v}$. This is precisely the reason for strong manifestation of the Dicke effect in the case of collisions preserving phase memory for the resonance in the line wing for low collision frequencies (for $\nu' \gg qv_T, kv_T|G|^2/\Omega^2$; however, ν' can be much smaller than $k_\mu v_T$). For $q \ll k$, the condition for manifestation of the Dicke effect, which was derived from the qualitative pattern, coincides with condition (26) weakly selective to the rates of interaction of atoms with radiation.

It can be seen from relations (36) that the Doppler shift in detuning $\Delta_\mu^{(+)}(\mathbf{v})$ in the particular case when $\mathbf{q} = -\mathbf{k}|G|^2/\Omega^2$ (we can substitute $\mathbf{k}_\mu$ for $\mathbf{k}$ in the case when $q \ll k$) disappears and all atoms resonantly interact with the probe field irrespective of their velocity. Thus, a resonance free of Doppler broadening emerges in this case in the vicinity of $\Omega_\mu = \Omega + |G|^2/\Omega$. This resonance is correctly described by formula (31) even in the absence of collisions if the condition $|\Omega|\Gamma \gg (k_\mu v_T)^2$ is satisfied (see formula (27)).

Figure 5b illustrate the behavior of the resonance free of Doppler broadening in the line wing (for $\mathbf{q} = -\mathbf{k}|G|^2/\Omega^2$) as a function of the collision frequency in the case when phase memory is preserved. The amplitude of the resonance is a complex function of the collision frequency. When the value of ν' increases from zero to a certain small value (to $\nu'/k_\mu v_T \sim 10^{-5}$ for the parameters corresponding to Fig. 5b), the resonance amplitude first decreases (curves *1*, *2*, and *3*) and then increases substantially with increasing ν' (curves *4* and *5*), attaining its maximum value for $\nu'/k_\mu v_T \sim 10^{-3}$. Upon a further increase in ν', the resonance amplitude decreases again (curves *6* and *7*).

4. CONCLUSIONS

In this paper, the spectrum of a weak probe field absorption by three-level atoms with the Λ configuration in the presence of a strong field at an adjacent transition is studied theoretically. It was assumed that atoms are in the buffer gas atmosphere and experience collisions with gas particles.

It is shown that most interesting features in the probe field spectrum are observed in the absence of collisional relaxation of low-frequency coherence at the transition between two lower levels of the Λ system. In this case, the absorption spectrum always exhibits clearly manifested anisotropy to mutual orientation of the wavevectors of the strong and probe radiations even in the limit of weak Doppler broadening relative to the collision frequency. Supernarrow resonances with a width much smaller than the natural width may appear; the characteristics of such resonances (width and amplitude) are connected in a certain manner with the diffusion coefficient for atoms interacting with radiation. This may form the basis for the development of a spectroscopic method for measuring the atomic diffusion coefficient.

The resonance in the far wing of the absorption line may experience Dicke collisional narrowing even for very small (as compared to the Doppler width) collision frequencies. Unexpectedly, it was found that the amplitude of the resonance in the far wing of the absorption line may exceed the amplitude of the resonance at the line center by several orders of magnitude. This result

is beyond the framework of conventional concepts, according to which the cross section of nonresonant radiative processes is always smaller than the cross section of resonant processes. Simple working equations are derived, which can be used for determining parameter $\tilde{\nu}'/\nu'$, viz., the degree of phase memory conservation in collisions at the *n–l* transition, from the relative amplitude of the resonance in the line wing.

The features of the probe field spectrum noted in this study are manifested most clearly for a high degree of phase memory conservation in collisions at the *n–l* transition (for $1 - \tilde{\nu}'/\nu' \ll 1$). For atoms of alkali metals (which are successfully simulated by the Λ level diagram) in the atmosphere of inert buffer gases, the cross section of collisional transitions between the *n* and *l* hyperfine structure components of the ground state is 6–10 orders of magnitude smaller than the gas-kinetic cross sections [11]. Consequently, we can expect for such objects a high degree of phase memory conservation in collisions, so that $1 - \tilde{\nu}'/\nu' \lesssim 10^{-6}$. Thus, atoms of alkali metals in the atmosphere of inert gases are suitable objects for detecting and studying the effects considered in this work.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project no. 04-02-16771), the program "Optical Spectroscopy and Frequency Standards" of the Physics Division of the Russian Academy of Sciences, and the program "Universities of Russia."

REFERENCES

1. S. G. Rautian, G. I. Smirnov, and A. M. Shalagin, *Nonlinear Resonances in Atomic and Molecular Spectra* (Nauka, Novosibirsk, 1979) [in Russian].
2. A. K. Popov, *Introduction in Nonlinear Spectroscopy* (Nauka, Novosibirsk, 1983) [in Russian].
3. V. S. Letokhov and V. P. Chebotaev, *High-Resolution Nonlinear Laser Spectroscopy* (Nauka, Moscow, 1990) [in Russian].
4. B. J. Feldman and M. S. Feld, Phys. Rev. A **5**, 899 (1972).
5. M. Kaivola, P. Thorsen, and O. Poulsen, Phys. Rev. A **32**, 207 (1985).
6. A. D. Wilson-Gordon, Phys. Rev. A **48**, 4639 (1993).
7. V. G. Arkhipkin, A. K. Popov, and A. S. Aleksandrovskiĭ, Pis'ma Zh. Éksp. Teor. Fiz. **59**, 371 (1994) [JETP Lett. **59**, 398 (1994)].
8. Yu. I. Belousov, E. V. Podivilov, M. G. Stepanov, and D. A. Shapiro, Zh. Éksp. Teor. Fiz. **118**, 328 (2000) [JETP **91**, 287 (2000)].
9. S. A. Babin, E. V. Podivilov, V. V. Potapov, *et al.*, Zh. Éksp. Teor. Fiz. **121**, 807 (2002) [JETP **94**, 694 (2002)].
10. Yu. I. Belousov and D. A. Shapiro, J. Phys. B: At. Mol. Opt. Phys. **36**, 1495 (2003).
11. W. Happer, Phys. Rep. **44**, 169 (1972).
12. B. D. Agap'ev, M. B. Gornyĭ, B. G. Matisov, and Yu. V. Rozhdestvenskiĭ, Usp. Fiz. Nauk **163** (9), 1 (1993) [Phys. Usp. **36**, 763 (1993)].
13. E. Arimondo, Prog. Opt. **35**, 257 (1996).
14. M. O. Skalli and M. S. Zubaĭri, *Quantum Optics* (Fizmatlit, Moscow, 2003) [in Russian].
15. O. Kocharovskaya, Phys. Rep. **219**, 175 (1992).
16. A. K. Popov, Izv. Ross. Akad. Nauk, Ser. Fiz. **60**, 99 (1996).
17. J. Mompart and R. Corbalan, J. Opt. B: Quantum Semiclassic. Opt. **2**, R7 (2000).
18. A. I. Parkhomenko and A. M. Shalagin, Zh. Éksp. Teor. Fiz. **120**, 830 (2001) [JETP **93**, 723 (2001)].
19. S. Chapman and T. G. Cowling, *Mathematical Theory of Non-Uniform Gases*, 2nd ed. (Cambridge Univ. Press, Cambridge, 1952; Inostrannaya Literatura, Moscow, 1960).
20. *Handbook of Mathematical Functions*, Ed. by M. Abramowitz and I. A. Stegun (Dover, New York, 1971; Nauka, Moscow, 1979).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 101, No. 6, 2005, pp. 999–1008.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 6, 2005, pp. 1145–1155.

ATOMS, MOLECULES, OPTICS

Spontaneous and Persistent Currents in Mesoscopic Aharonov–Bohm Loops: Static Properties and Coherent Dynamic Behavior in Crossed Electric and Magnetic Fields¶

I. O. Kulik

Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY, 11794-3800 USA
Department of Physics, Bilkent University, Ankara, 06533 Turkey
e-mail: iokulik@yahoo.com

Received June 8, 2004

Abstract—Mesoscopic or macromolecular conducting rings with a fixed number of electrons are shown to support persistent currents due to the Aharonov–Bohm flux, and the "spontaneous" persistent currents without the flux when structural transformation in the ring is blocked by strong coupling to the externally azimuthal-symmetric environment. In the free-standing macromolecular ring, symmetry breaking removes the azimuthal periodicity, which is further restored at the increasing field, however. The dynamics of the Aharonov–Bohm loop in crossed electric and magnetic fields is investigated within the tight-binding approximation; we show that transitions between discrete quantum states occur when static voltage pulses of prescribed duration are applied to the loop. In particular, the three-site ring with one or three electrons is an interesting quantum system that can serve as a qubit (quantum bit of information) and a qugate (quantum logical gate) because in the presence of an externally applied static electric field perpendicular to a magnetic field, the macromolecular ring switches between degenerate ground states mimicking the NOT and Hadamard gates of quantum computers. © *2005 Pleiades Publishing, Inc.*

1. PERSISTENT CURRENTS IN MESOSCOPIC SYSTEMS

Persistent currents have been predicted for mesoscopic conducting ballistic or quasiballistic loops ([1][1] and references therein, [2]) that do not show the effect of superconductivity and that have been extended to diffusive rings [3]. The current appears in the presence of a magnetic field as a result of the Aharonov–Bohm effect [4], demonstrating the special role of the vector potential in quantum mechanics. As discussed in [5], persistent currents are similar to orbital currents in normal metals first considered by Teller [6] in his interpretation of Landau diamagnetism in metals [7], but are specific to the doubly connected geometry of conductors (loops, hollow cylinders, etc.). Persistent currents have been observed in indirect [8, 9] as well as direct [10, 11] experiments, showing the single-flux-quantum $\Phi_0 = hc/e$ periodicity in the resistance of thin Nb wires [8] and networks of isolated Cu rings [9], and in single-loop experiments on metals [10] and semiconductors [11]. In [12], the periodic variation of resistivity in molecular conducting cylinders (carbon nanotubes) was attributed to the Altshuler–Aronov–Spivak effect [13], a companion to the classical Aharonov–Bohm mechanism with the twice smaller periodicity in magnetic flux $\Phi_1 = hc/2e$. A further trend in macromolecular persistent currents [14–16] is in the quantum computational [17] prospects of using the Aharonov–Bohm loops as qubits with an advantage of easier (radiation-free) manipulation of qubit states, and in the increased decoherence times compared to macroscopic "Schrödinger cat" structures (Josephson junctions).

The present paper focuses on ballistic Aharonov–Bohm rings, like those naturally found in molecular crystals with metalloorganic complexes as the building blocks [18, 19]. We approximate such macromolecular structures as rings with resonant hopping of electrons between the near-site atoms or complexes serving as electron localization sites. As shown in [14], the smallest (three-site) persistent current ring displays a Λ-shaped energy configuration (Fig. 1) with two degenerate ground states, at the external flux through the ring equal to half the normal-metal flux quantum, $\Phi = hc/2e$. At a certain number of electrons in the ring, persistent current appears at zero field (the "spontaneous" current). The spontaneous persistent current loop, to be discussed below, achieves the degenerate state at zero

¶ The text was submitted by the author in English.

[1] This paper proved exact periodicity of ring energy as a function of the magnetic flux with the period hc/e, although with an indefinite amplitude.

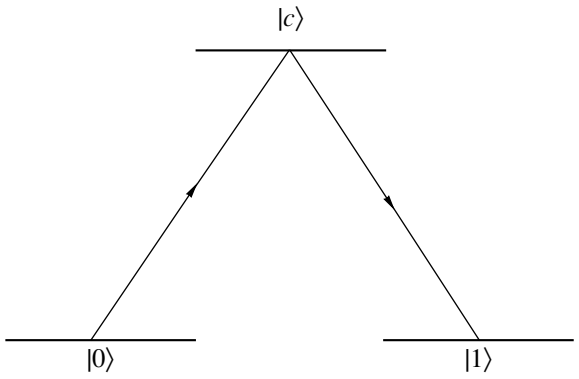


Fig. 1. A Λ-shaped energy configuration in the Aharonov–Bohm ring. Arrows indicate a transition between degenerate states $|0\rangle$ and $|1\rangle$ through virtual transition to the control state $|c\rangle$.

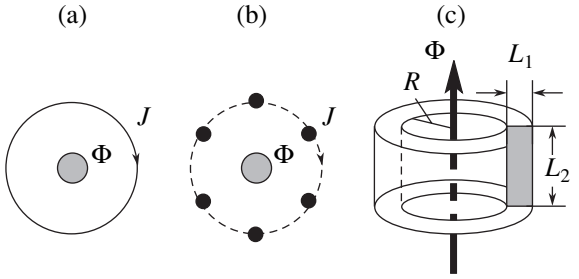


Fig. 2. (a) Models of mesoscopic and nanoscopic Aharonov–Bohm loops: a one-dimensional continuous loop; (b) a discrete loop with regularly spaced centers of electron localization (sites); (c) a 3-dimensional loop in the form of a cylinder with a longitudinal dimension of $L = 2\pi R$ and transverse dimensions of L_1, L_2.

field or, if the degeneracy is lifted by the electron–phonon coupling, at reasonably small fields.

Persistent current is a voltage-free nondecaying current that exists as a manifestation of the fact that the ground state of a doubly connected conductor in a magnetic field is a current-carrying one. This statement was proved for ballistic loops [2] and for diffusive rings [3]. There is no principal difference between these extremes. Counterintuitively, a ballistic structure does not show infinite conductivity, as was sometimes naively supposed; a dc resistance of the loop is infinite rather than zero when a dc electric field is applied to the system. In the case where a current is fed through the structure, no voltage appears provided the magnitude of the current is smaller than a certain critical value. This applies to both elastic and inelastic scatterings. The magnitude of the critical current of a ballistic ring smoothly matches the current of the diffusive ring when the mean free path of the electron becomes large. In a dirty limit, $l \ll L$, where l is the electron mean free path and L is the ring circumference, the critical value of the persistent current decreases proportionally to l/L according to [20], or to $(l/L)^{1/2}$ according to numerical simulation [5]. The nondecaying current does not even require severe restrictions on the so-called "phase breaking" mean free path l_φ. In fact, the normal-metal supercurrent is an analogue of the "nonquantum" Josephson effect [21, 22], the one in which the phase of superconductor is considered a classical variable. Stronger criteria (the dephasing length larger than the system size, and the analogous requirement in the time domain, that the "decoherence time" is larger than the characteristic time of observation) apply to persistent current rings as quantum computational tools, which are analogs of macroscopic quantum tunneling [23–26]. Persistent current is a thermodynamic property, clearly distinct from the dissipative currents in conductors, and can in principle exist in a system that has the vanishing Ohmic conductance.

2. SPONTANEOUS PERSISTENT CURRENTS

Persistent current in a ballistic ring appears due to the Aharonov–Bohm flux. The current, however, can also occur when the external magnetic field is zero, in which case it is called the spontaneous current. Such a situation was noticed accidentally by various authors, in particular, [27, 28], but did not seem convincing, did not attract attention due to fixed chemical potential configuration studied, and was attributed to the effect of Peierls instability in the ring [29–32] (with the latter paper criticized [33, 34] in regard to the inaccuracy of the mean-field approximation). In fact, the fixed-number-of-particle ring with an odd number of electrons displays a number of structural instabilities, of which the Peierls transformation [35] and the Jahn–Teller effect [36] are the best known examples, or generally, a more complex atom rearrangement when the ground state proves degenerate in a symmetric configuration.

The origin of the spontaneous current can be understood as follows. We consider a one-dimensional ring in the field of a vector potential created by a thin, infinitely long solenoid perpendicular to the plane of the ring and piercing the ring (Fig. 2a). The electron energy in the ring is

$$\varepsilon_n = \frac{h^2}{2mL^2}\left(n - \frac{LA}{\Phi_0}\right)^2, \qquad (1)$$

where $A = \Phi/L$ is the angular component of the vector potential (Φ is the total magnetic flux of the solenoid) and $n = 0, \pm 1, \pm 2, \ldots$. Such a state corresponds to the current

$$J = -\frac{c\partial\varepsilon_n}{\partial\Phi} = \frac{eh}{mL^2}\left(n - \frac{\Phi}{\Phi_0}\right),$$

which is zero at $\Phi = 0$ and $n = 0$, but is nonzero at $n \neq 0$ even at zero flux. At $T = 0$, electrons, in the total number N, occupy the lowest possible energies compatible with

the Pauli exclusion principle, i.e., such that each state is occupied with two electrons with opposite spins at most. Therefore, the ground state of one or two electrons is that of $n = 0$, and hence has zero current at $\Phi = 0$. But the state with the next electron number, $N = 3$, already resumes at $n = 1$ or $n = -1$, or is in a superposition of these states, $\alpha|1\rangle + \beta|-1\rangle$, depending on the way the state at the initial condition is prepared, and therefore carries a current unless $\alpha \neq \beta$. If there is no decoherence of the state due to the interaction of the loop with the environment, the current persists in time without any voltage applied along the loop. This applies to a ballistic perfectly symmetric ring. The inhomogeneity in the ring, as well as scattering of electrons by impurities, may result in a nondegenerate current-free state. This is illustrated in Fig. 3 for the ring with a δ-functional barrier $V_0\delta(x)$, which results in the Kronig–Penney equation for energy,

$$\cos(2\pi k) + \frac{V_0 L}{2\varepsilon_0}\frac{\sin(2\pi k)}{2\pi k} = \cos\left(2\pi\frac{\Phi}{\Phi_0}\right). \qquad (2)$$

The electron energy is $\varepsilon = \varepsilon_0 k^2$, where $k = k_n$ is one of solutions to Eq. (2) and $\varepsilon_0 = h^2/2mL^2$. The same conclusion is obtained for a discrete Aharonov–Bohm ring (Fig. 2b), to be considered in detail below.

Figure 4 shows the maximum value of persistent current, as well as that of the spontaneous current introduced above, versus the number of electrons in a three-dimensional ballistic ring (the one with the electron mean free path $l = \infty$) modeled as a finite-length hollow cylinder (Fig. 2c) with the rectangular cross section $L_1 \times L_2$ containing a finite number of perpendicular electron channels

$$N_\perp = \frac{L_1 L_2 k_F^2}{2\pi^2}.$$

We note that the magnitude of the current in a ballistic ring is not ev_F/L, as is sometimes suggested (v_F is the Fermi velocity), but

$$J_{max} \sim \frac{ev_F}{L} N_\perp^{1/2}$$

(see also [2]). The dependence $J_{max}(N)$ at $T = 0$ is irregular due to the contribution to the total current of both the negative and positive terms originating from different electron eigenstates.

Figure 5 explains the origin of persistent current as a bistability effect in a ring. While the electron energy has a minimum at $\Phi = 0$ for an even number of electrons, it acquires a maximum when the number of electrons is odd. The inductive energy, to be included below, shifts the position of minima in that curve only very slightly. The spontaneous current has the same order of magnitude as the maximum persistent current, and it is an inseparable part of the Aharonov–Bohm effect in a ballistic ring.

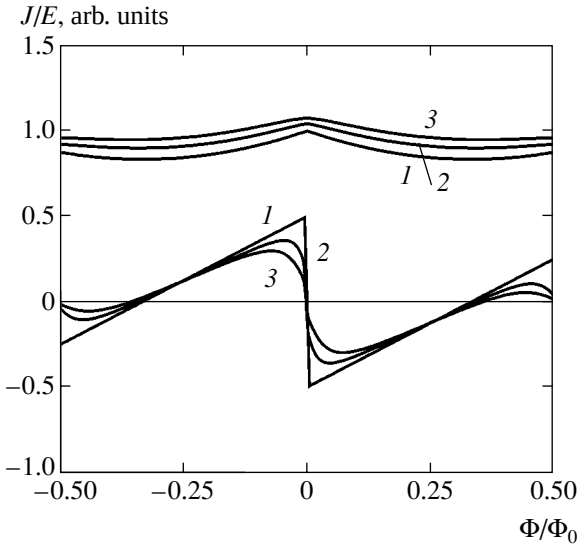


Fig. 3. Ground state energies and currents in the continuous ring with 3 electrons at various strengths of the barrier: g = 0 (*1*), 1 (*2*), 2 (*3*).

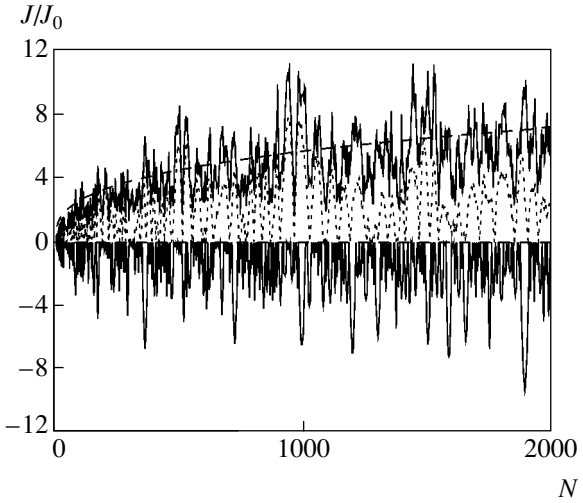


Fig. 4. Persistent current versus the number of electrons in a ring with ratio cross-sectional dimensions $L : L_1 : L_2$ = 10 : 1 : 1 (configuration with spin). The upper curve is the maximum current in units of $J_0 = ev_F/L$ at given N; the dotted curve is the amplitude of the first harmonic of $J_{pers}(\Phi)$; and the curve at negative J is the spontaneous persistent current, also in units of J_0. The dashed curve is the square root of the number of perpendicular channels $N_\perp$.

In a one-dimensional loop, discrete quantum states are

$$\psi_n = \frac{1}{\sqrt{L}} e^{in\theta}, \qquad (3)$$

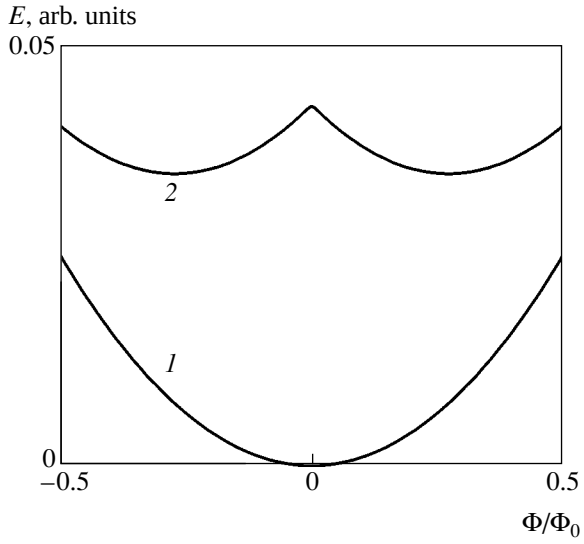


Fig. 5. Examples of the occurrence of a bistable configuration in a ring. Energy versus flux in a ring of 10 (*1*) and 11 (*2*) electrons. Curve *2* is shifted downward for convenience but is not reset.

where θ is the azimuthal angle, with the energies given by (1) plus the inductive energy of the current. For the loop with three electrons, this gives the total energy,

$$E(f) = \varepsilon_0\left[f^2 + \frac{1}{2}(\pm 1 - f)^2\right] + \frac{\mathcal{L}J_0^2}{2c^2}j^2(f), \qquad (4)$$

corresponding at $\Phi = 0$ to two spin-1/2 states with $n = 0$ and one state with $n = 1$ or $n = -1$. The last term in Eq. (4) is the magnetic inductive energy and $\mathcal{L}$ is the inductance (of the order of the ring circumference, in the units adopted). The current

$$J = -\frac{e}{h}\frac{\partial E}{\partial f}$$

is equal to

$$J(f) = J_0(\pm 1 - 3|f|), \quad J_0 = \frac{e\varepsilon_0}{h} \qquad (5)$$

and is nonzero at $f = 0$ in either of the states ±. The ratio of the magnetic energy to the kinetic energy is on the order of

$$\eta = \frac{\mathcal{L}J_0^2}{2c^2\varepsilon_0} \approx \frac{e^2}{4\pi mc^2 R} \sim 10^{-6}\frac{a_0}{R}, \qquad (6)$$

where a_0 is the Bohr radius. This is a very small quantity, and therefore the magnetic energy is unimportant in the energy balance of the loop. The total flux in the loop is $f = f_{\text{ext}} + 2\eta j(f)$, where f_{ext} is the external flux and $j(f) = J(f)/J_0$. The correction to the externally applied flux is significant only at very small fields $f_{\text{ext}} \sim \eta$; otherwise we can ignore this contribution.

When a persistent-current loop is placed in an electric field perpendicular to a magnetic field, the system coherently switches between the discreet states of the loop providing for quantum transitions (quantum logical gates) in the loop performing as a qubit in a quantum computer. This aspect of persistent currents in ballistic loops is analyzed in Section 3.

The property of a nonzero spontaneous persistent current thus demonstrated for noninteracting electrons survives strong electron–electron coupling but collapses when the coupling to the lattice is included. This is considered in detail in Section 4. In what follows, the structural transformation in the ballistic ring is investigated in an exact way by considering the ring dynamics in the tight-binding approximation [37, 38]. The "lattice" (the atomic configuration of the loop) can respond to the bistable state by a readjustment of atoms similar to the Peierls transition (doubling of the lattice period in a one-dimensional atomic chain, see, e.g., [39, 40]), or by a more general lattice transformation that does not reduce to simple doubling. When the loop is in the rigid background in the periodic lattice on a substrate of a much stronger bound solid, the degeneracy may not be lifted, or may remain in a very narrow interval of the externally applied field.

3. DYNAMICS OF PERSISTENT CURRENTS IN CROSSED ELECTRIC AND MAGNETIC FIELDS

The Hamiltonian of the ring consisting of N sites localizing electrons at equidistant angular positions is $\theta_n = 2\pi n/N$ is

$$H_0 = -\tau\sum_{n=1}^{N}(a_n^+ a_{n+1}e^{i\alpha} + a_{n+1}^+ a_n e^{-i\alpha}), \qquad (7)$$

where a_n^+ is a fermionic operator creating (and a_n, annihilating) the electron at the site $\mathbf{R}_n$ in the ring with the periodic boundary condition $a_{N+1} = a_1$, and $\alpha = 2\pi\Phi/N\Phi_0$ is the phase related to the Aharonov–Bohm flux threading the ring. Placing the ring in the homogeneous electric field perpendicular to the magnetic field (Fig. 6) results in the extra term

$$H_1 = V_0\sum_{n=1}^{N}\cos\frac{2\pi n}{N}a_n^+ a_n \qquad (8)$$

being added to the Hamiltonian. The Hamiltonian H_0 is diagonalized by the angular momentum (i.e., $m = 0, 1,$

…, $N-1$) eigenstates $A_m^+|0\rangle$ such that

$$A_m^+ = \frac{1}{\sqrt{N}}\sum_{n=0}^{N-1} a_n^+ \exp\frac{2\pi imn}{N}. \quad (9)$$

These states have the energies

$$\varepsilon_m = -2\tau\cos\frac{2\pi}{N}\left(m - \frac{\Phi}{\Phi_0}\right) \quad (10)$$

plotted versus the flux in Fig. 7. The electronic configuration at $\Phi = \Phi_0/2$ has a Λ-shaped energy structure with two degenerate ground states shown in Fig. 1, which were suggested as $|0\rangle$ and $|1\rangle$ components of a qubit in [14, 15]. The time evolution of angular-momentum eigenstates $A_m^+|0\rangle$ is periodic at certain values of V_0 and at the value of the flux equal to half the flux quantum $\Phi_0/2 = hc/2e$.

In the eigenbasis of the operators A_m, the Hamiltonian $H_0 + H_1$ at $N = 3$ in the absence of an electric field is transformed into the diagonal form (we scale all energies in units of τ)

$$H_0 = \sum_m \varepsilon_m A_m^+ A_m = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad (11)$$

and the Hamiltonian H_1 becomes

$$H_1 = \begin{pmatrix} 0 & v & v \\ v & 0 & v \\ v & v & 0 \end{pmatrix}, \quad (12)$$

where $v = V_0/2\tau$. We let the $m = 1$ and $m = 3$ states be denoted by $|0\rangle$ and $|1\rangle$, in the qubit terminology, and the excited state $m = 2$ by $|c\rangle$ (the "control" state coupling qubit states to the "qugate," or the quantum logic gate).

The eigenstates of $H_0 + H_1$ versus v at $\Phi = \Phi_0/2$ are presented in Fig. 8. We assume that at $t \le 0$, the potential is $V_0 = 0$, such that the system at $t = 0$ is a superposition of the angular momentum states $A_m^+|0\rangle$ with certain amplitudes $C_m(0)$. At a later time and at a constant value of V_0, $C_n(t)$ evolves as

$$C_n(t) = \sum_m \exp(-i(H_0 + H_1)t)_{mn} C_m(0). \quad (13)$$

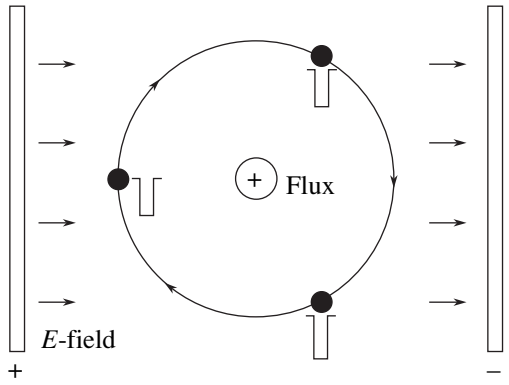


Fig. 6. Scheme of a 3-site qubit in the electric field perpendicular to the magnetic field.

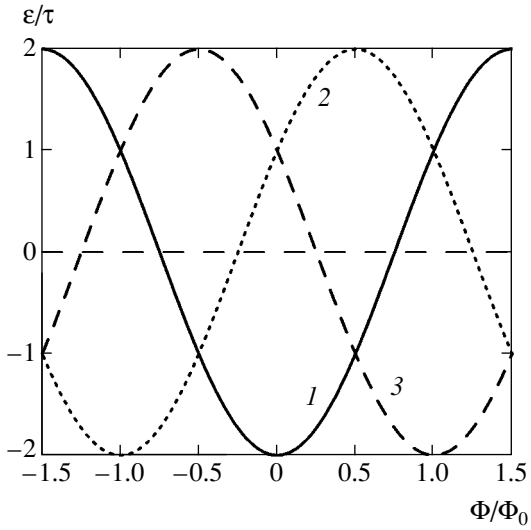


Fig. 7. Curves *1* and *3* are energy versus magnetic flux dependences in the degenerate states carrying opposite currents. The current is found as the derivative $j = -c\partial\varepsilon/\partial\Phi$. Curve *2* corresponds to the zero-current virtual state at the operating point of a qubit at the half-flux quantum $\Phi = \Phi_0/2$.

For a step function $V(t) = V_0\theta(t)$, this gives the dependence [14]

$$C_n(t) = \sum_{m,k} S_{kn}^{-1}(V_0)\exp(-iE_k t)S_{mk}(V_0)C_m(0), \quad (14)$$

where $\varepsilon_k(V_0)$ are eigenenergies of the Hamiltonian $H_0 + H_1(V_0)$ and $S_{nm}(V_0)$ are the unitary matrices transforming from the noninteracting eigenbasis (the one corresponding to H_0) to the eigenbasis of the full Hamiltonian $H_0 + H_1$. It is implied in Eq. (14) that at a fixed value of V_0, the time evolution is performed as the interplay between the three different eigenenergies. This is sufficient evidence that if the eigenenergies are appropriately adjusted, the population of the auxiliary state

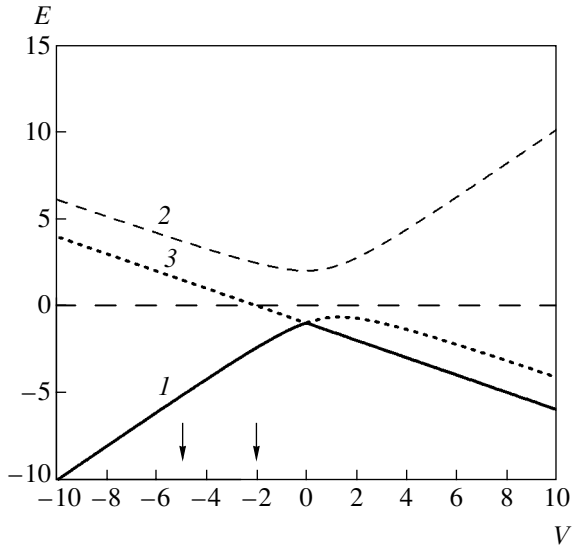


Fig. 8. Energy versus electrostatic potential. Curves *1* and *3* (solid and dotted lines) are the energies that become degenerate at $V_0 = 0$, and curve *2* (the dashed line) is the energy of the auxiliary control state $|c\rangle$. The arrows indicate the values of the potential V_0 corresponding to the operational points of the bit-flip and Hadamard gates.

(in the angular-momentum basis) can vanish for certain initial conditions. At these time instants, the three-state system instantaneously collapses into the qubit subspace without loss of any information if the auxiliary state $|c\rangle$ was initially unoccupied. A necessary condition for the instantaneous collapse into the qubit subspace (i.e., the degenerate-level subspace) is a commensuration condition between the eigenenergies $\varepsilon_k(V_0)$, $k = 1, 2, 3$ such that the exponential factors in Eq. (14) destructively interfere at fixed tune instants to destroy the nondiagonal correlations. The required commensuration can be expressed by the condition

$$\varepsilon_3 - \varepsilon_1 = \nu(\varepsilon_2 - \varepsilon_3) \tag{15}$$

for integer ν. Equation (15) guarantees periodic collapses of the wavefunction onto the desired basis, and the next step is to find whether the desired qugate operations can be realized simultaneously in this basis. For the corresponding values of the potential respecting Eq. (15), we find

$$V_0(\nu) = -\frac{2}{3\nu}[\nu^2 + \nu + 1 + (\nu - 1)\sqrt{\nu^2 + 4\nu + 1}]. \tag{16}$$

In particular, we note that for $\nu = 1$, we have $V_0^{(1)} = -2$ and at $\nu = 3$, we have

$$V_0^{(3)} = -\frac{2}{9}(13 + 2\sqrt{22}) = -4.9735,$$

and we succeeded in finding two qugates in our first few attempts. As shown below, these two cases yield the bit-flip and Hadamard transformations of the qubit [17].

The $\nu = 1$ case can be explicitly proved by verifying the identity

$$\exp\left\{-it\begin{pmatrix} -1 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & -1 \end{pmatrix}\right\} = \frac{1}{2}\begin{pmatrix} 1+c+s & s & -1+c+s \\ s & 2(c-s) & s \\ -1+c+s & s & 1+c+s \end{pmatrix}, \tag{17}$$

where

$$c = \cos(t\sqrt{6}), \quad s = i\sqrt{\frac{2}{3}}\sin(t\sqrt{6}).$$

At $s = 0$ (i.e., $c = \pm 1$), the transformation matrix of qubit states is block-diagonalized in the subspace of states 1, 3 (i.e., the qubit states $|0\rangle, |1\rangle$) and the upper state 2 (i.e., the auxiliary "control" state $|c\rangle$). In particular, for $c = -1$, the bit-flip is performed between the qubit states.

In Fig. 9, the populations $p_n(t) = |C_n(t)|^2$ of the states are plotted for the mentioned cases $\nu = 1$ and $\nu = 3$. The instantaneous collapse to the qubit subspace is obtained at $t = t_1$ for $\nu = 1$ and at $t = t_3$ for $\nu = 3$ if the auxiliary level is unoccupied at $t = 0$. We found these critical times as (in units of $\hbar/\tau$)

$$t_1 = \frac{\pi}{\sqrt{6}} = 1.2825, \qquad t_3 = \frac{\pi}{2[E_2(V_0) - E_3(V_0)]_{\nu=3}} = 0.7043, \tag{18}$$

where the eigenenergies are

$$E_{1,3}(V_0) = \frac{1 + V_0/2}{2} \mp \frac{3}{2}\sqrt{1 - \frac{V_0}{2} + \frac{V_0^2}{4}}, \qquad E_2(V_0) = -1 - \frac{V_0}{2} \tag{19}$$

for $V_0 \le 0$. We note that the configuration (t_1, $\nu = 1$) performs the bit-flip $|0\rangle \longleftrightarrow |1\rangle$, whereas ($t_3$, $\nu = 3$) creates the equally populated Hadamard-like superpositions of $|0\rangle$ and $|1\rangle$. These operations are represented in the qubit subspace by the matrices (overall phases are not shown)

$$G_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ and } G_3 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -i \\ -i & 1 \end{pmatrix}. \tag{20}$$

The dotted lines show the time dependence of the auxiliary population. The arrows indicate the "operational

point" of the qugate, the time of evolution corresponding to the return to the invariant qubit. The G_1 transformation manifests the bit-flip (NOT gate) and G_3 is similar to the Hadamard gate [17] except for the phase shift $\pi/2$.

4. QUANTUM BISTABILITY AND SPONTANEOUS CURRENTS IN A COUPLED ELECTRON–PHONON SYSTEM

In the tight-binding approximation, the Hamiltonian of the loop in the secondary quantized form is given by

$$H = \sum_{j=1}^{N} (\tau_j a_{j\sigma}^{+} a_{j+1,\sigma} e^{i\alpha_j} + \text{H.c.}) + U \sum_{j=1}^{N} n_{j\uparrow} n_{j\downarrow}$$

$$+ V \sum_{j=1,\sigma,\sigma'}^{N} n_{j\sigma} n_{j+1,\sigma'} \qquad (21)$$

$$+ \frac{1}{2} W \sum_{j=1}^{N} (\theta_j - \theta_j^0)^2 + \frac{1}{2} K \sum_{j=1}^{N} (\theta_j - \theta_{j+1})^2,$$

where τ_j is the hopping amplitude between two adjacent configurational sites, j and $j+1$,

$$\tau_j = \tau_0 + g(\theta_j - \theta_{j+1}), \quad n_{i\sigma} = a_{i\sigma}^{+} a_{i\sigma}, \qquad (22)$$

and

$$\alpha_j = \frac{2\pi f}{N} + (\theta_j - \theta_{j+1}) f \qquad (23)$$

is the Aharonov–Bohm phase (a Peierls substitution for the phase of hopping amplitude). Next, $a_{j\sigma}^{+}$ is the creation (and $a_{j\sigma}$ is the annihilation) operator of the electron at site j with spin σ; θ_j, $j = 1, 2, \ldots, N$ are the angles of distortion of site locations from their equilibrium positions $\theta_j^0 = 2\pi j/N$ satisfying the requirement

$$\sum_{j=1}^{N} \theta_j = 0;$$

and g is the electron–phonon coupling constant. The interaction in Eq. (22) reflects the property that the hopping amplitude depends on the distance between the localization positions and assumes that the displacement $\theta_j - \theta_{j+1}$ is small in comparison to $2\pi/N$. U and V are Hubbard parameters of the on-site and intrasite interactions. W is the binding energy of the loop to external environment (a substrate) such that the loop passes into the azimuthally symmetric configuration $\theta_i = \theta_i^0$ as $W \longrightarrow \infty$.

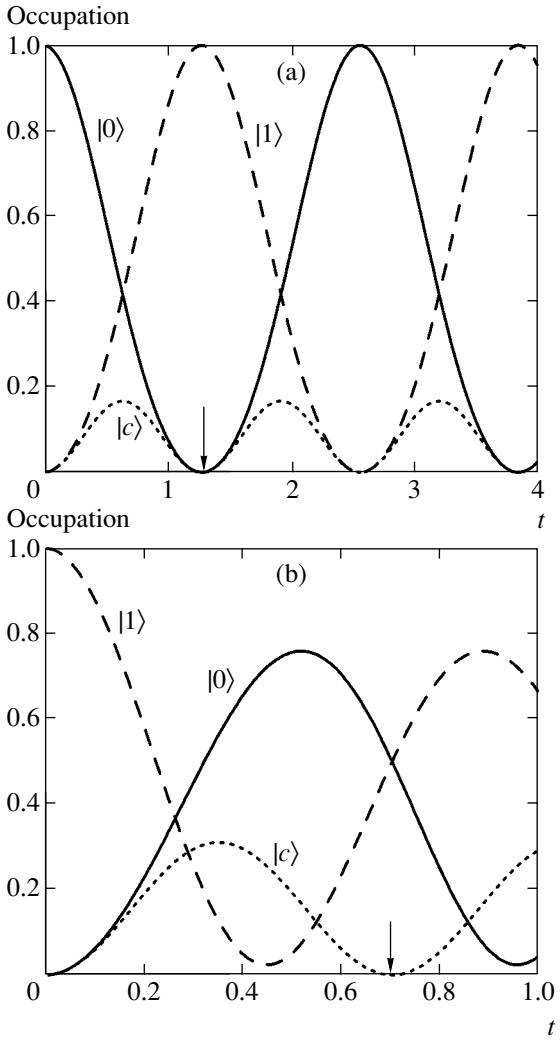


Fig. 9. Evolution diagrams of the quantum gate G_1 (a) and G_3 (b). Solid and dashed lines are the time dependences of the population of states $|0\rangle$ and $|1\rangle$. The dotted line shows the time dependence of the auxiliary-state population. The arrow indicates the "operational point" of the qugate, i.e., the evolution time corresponding to the return to the invariant qubit subspace.

The parameters are assumed such that the system is not superconductive (e.g., $U > 0$; anyway, the superconductivity is not allowed for a 1D-system and it is forbidden for a small system). The last term in Hamiltonian (21) is the elastic energy and K is the stiffness parameter of the lattice.

In the smallest loop, the one with three sites ($N = 3$), only two free parameters of the lattice displacement, X_1 and X_2, remain:

$$\theta_1 = X_1 + X_2, \quad \theta_2 = -X_1 + X_2, \quad \theta_3 = -2X_2, \qquad (24)$$

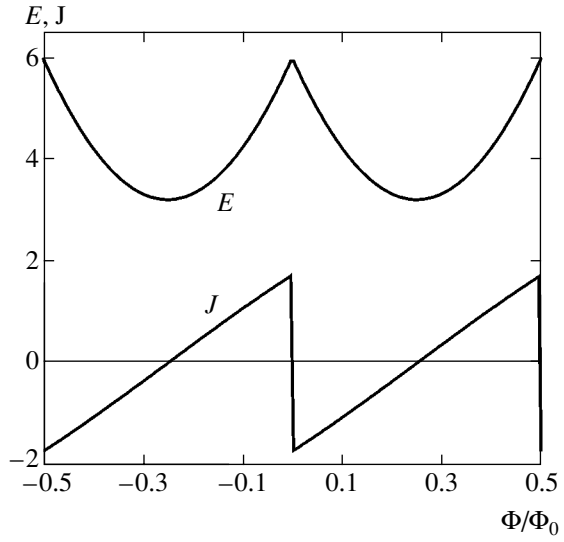


Fig. 10. Lower curve: current versus magnetic flux in a 3-site loop with 3 noninteracting electrons. Upper curve: energy versus flux in the loop. The hopping parameter is $\tau_0 = -1$. The energy is reset and arbitrarily shifted upward for clarity.

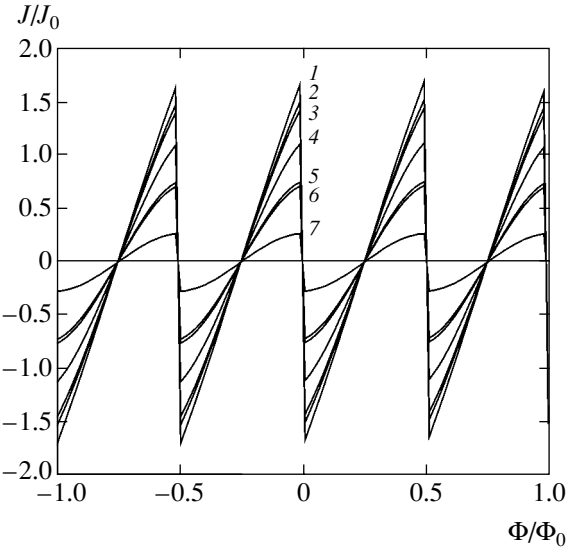


Fig. 11. Spontaneous persistent current versus flux for $\tau_0 = -1$ and various values of the Hubbard parameter U: $U = 0$ (*1*), –2 (*2*), 2 (*3*), –5 (*4*), 5 (*5*), –10 (*6*), 10 (*7*).

which are decomposed with respect to secondary quantized Bose operators b_1 and b_2 as

$$X_1 = \left(\frac{3K}{\omega}\right)^{1/4}(b_1 + b_1^+),$$
$$X_2 = 3\left(\frac{K}{3\omega}\right)^{1/4}(b_2 + b_2^+). \tag{25}$$

System (21) is solved numerically with the ABC compiler [41], which includes the creation–annihilation operators as its parameter types. These are generated as compiler macros with sparse matrices

$$A_n = C_n^{(N_1)} \otimes 1^{(N_2)}, \quad \text{fermionic sector},$$
$$B_n = 1^{(N_1)} \otimes C_n^{(N_2)}, \quad \text{bosonic sector}, \tag{26}$$

where $1^{(N)}$ is the unit matrix of size 2^N; $C_n^{(N)}$, $n = 1, \ldots, N$ are Fermi/Bose operators in a space of the same dimension,

$$C_n^{(N)} = (u\otimes)^{N-n} a (\otimes v)^{n-1}; \tag{27}$$

and a, u, and v are the 2×2 matrices (with $\otimes$ being the symbol of the Kronecker matrix product):

$$a = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},$$
$$u = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad v = \begin{pmatrix} 1 & 0 \\ 0 & \eta \end{pmatrix}, \tag{28}$$

and

$$\eta = \begin{cases} -1, & \text{fermionic sector}, \\ 1, & \text{bosonic sector}. \end{cases} \tag{29}$$

The bosons are considered hardcore bosons, such that there are only two discrete states for each mode of displacement. We calculate the ground state of Hamiltonian (21) as a function of the magnetic flux f (a classical variable). In application to real atomic (macromolecular) systems, we can consider X_1 and X_2 as classical variables because quantum uncertainties in the coordinates ($\Delta X_{1,2} \sim (\hbar/M\omega)^{1/2}$) are typically much smaller than the interatomic distances (M is the mass of an atom and $\omega \sim 10^{13}$ s^{-1} is the characteristic vibration frequency). The energy of the loop is calculated as a function of X_1 and X_2 and further minimized with respect to X_1 and X_2 for each value of f. The nonzero values of X_1 and X_2 signify a "lattice" (the ionic core of the macromolecule) instability against the structural transformation, analogous to the Peierls transition.

In the noninteracting system (U, V, W, $g = 0$), the energy versus the flux f shows a kink with a maximum at $f = 0$ (Fig. 10) in the half-filling case, i.e., at a number of electrons n equal to the number of sites N, as well as in a broader range of values of n at larger N. Actually, as is clear from Fig. 4, such a dependence is typical of any $N \geq 3$ system for a number of (fixed) values of n.

The 3-site loop's $E(f)$ dependence is shown in Fig. 10 together with the dependence of the current on f. The latter shows a discontinuity at $f = 0$ of the same order of magnitude as the standard value of the

persistent current. The current at $f = 0$ is paramagnetic because the energy vs. flux has a maximum rather than a minimum at $f = 0$. The on-site interaction reduces the persistent current amplitude near zero flux (Fig. 11) but does not remove its discontinuity at $f = 0$. Therefore, the strongest opponent of the Aharonov–Bohm effect, the electron–electron interaction, leaves the current qualitatively unchanged.

On the other hand, the electron–phonon interaction (considered here classically in regard to lattice vibration) flattens the $E(f)$ dependence near the peak value (see Fig. 12a). At large stiffnesses K, this flattening remains important only for small magnetic fluxes, much smaller than the flux quantization period $\Delta\Phi = \Phi_0$. We note that the persistent current peak reduces in its amplitude only slightly near $\Phi = 0$. As is seen from Fig. 12b, the electron–phonon interaction splits the singularity at $\Phi = 0$ to two singularities at $\Phi = \pm\Phi_{sing}$. Outside the interval $-\Phi_{sing} < \Phi < \Phi_{sing}$, the structural transformation is blocked by the Aharonov–Bohm flux. The range of magnetic fluxes between $-\Phi_{sing}$ and Φ_{sing} determines the domain of the developing lattice transformation, which signifies itself with nonzero values of lattice deformations X_1 and X_2. This property allows us to suggest that the spontaneous persistent current state (a peak of dissipationless charge transport at or near the zero flux) remains at a nonzero Φ when the electron–phonon coupling is not too strong or when the lattice stiffness is larger than a certain critical value.

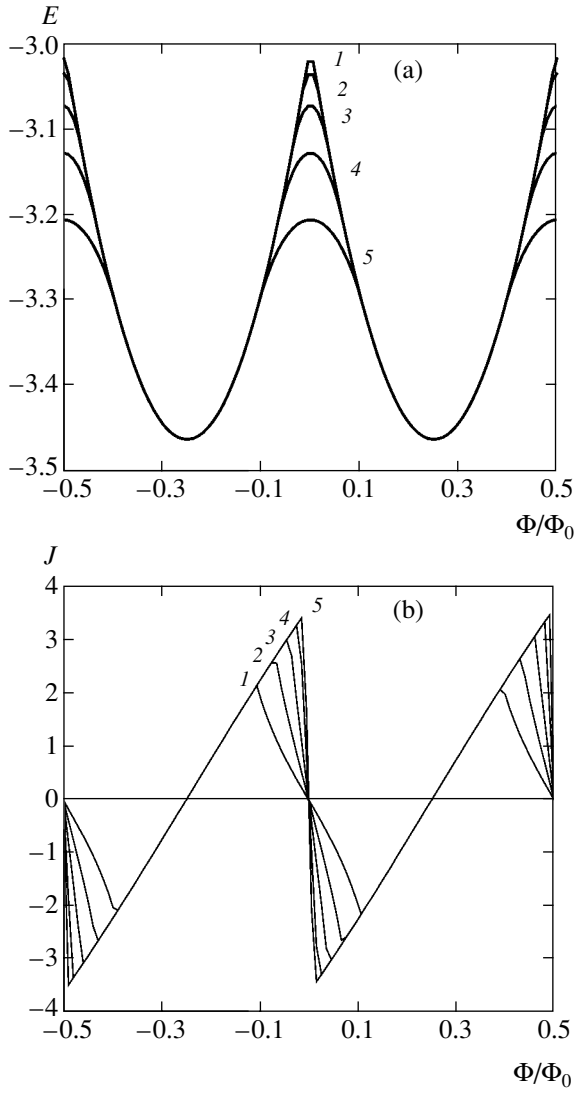


Fig. 12. Energy (a) and current (b) versus flux in a loop of noninteracting electrons coupled to the lattice with the coupling parameter $g = 1$ and various values of the stiffness parameter K: $K = 2$ (*1*), 3 (*2*), 5 (*3*), 10 (*4*), 20 (*5*).

5. DISCUSSION

In conclusion, we considered the Aharonov–Bohm effect in an angular-periodic macromolecular structure, like that of an aromatic cyclic molecule, and established the existence of a persistent current and also a spontaneous current when the Aharonov–Bohm flux is not applied to the ring. Strong coupling of electron hopping to the ion core of the molecule removes the spontaneous current, which is nevertheless restored at a (small) magnetic field, or when the loop has large stiffness or is strongly bound to an external azimuthal-periodic environment (a substrate). Degenerate states of the loop at $\Phi = \Phi_0/2$ and at $\Phi = 0$ may serve as components of a qubit that are operated by static voltages applied in the plane of the loop perpendicular to the direction of the Aharonov–Bohm flux.

The papers of Gatteschi *et al.* [18, 19] are particularly noteworthy, in which an azimuthal-periodic molecular structure (a "ferric wheel" $[Fe(OMe)_2(O_2CCH_2Cl)]_{10}$) exhibited periodic variation of its magnetization as a function of the magnetic flux; we assume that the periodicity with large period can be attributed to persistent currents. The above macromolecular structure is more complex than the one we considered because it contains magnetic ions with strong exchange interactions such that the actual magnetic field in the ring may be larger than the externally applied field. If this suggestion proves correct, it will open the possibility of engineering macromolecular structures (qubits and qugates) based on the Aharonov–Bohm effect, for purposes of quantum computation. Apart from this, the very existence of a nonzero nondecaying current in a nonsuperconductive system is, in our opinion, of fundamental physical interest.

ACKNOWLEDGMENTS

I would like to thank Prof. D. Averin for helpful discussions and advice, and Prof. K. Likharev for comments on quantum computational aspects of nanoscale physics.

REFERENCES

1. F. Bloch, Phys. Rev. B **2**, 109 (1970).
2. I. O. Kulik, Pis'ma Zh. Éksp. Teor. Fiz. **11**, 407 (1970) [JETP Lett. **11**, 275 (1970)].
3. M. Büttiker, Y. Imry, and R. Landauer, Phys. Lett. **96A**, 365 (1983).
4. Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959).
5. I. O. Kulik, in *Quantum Mesoscopic Phenomena and Mesoscopic Devices in Microelectronics*, Ed. by I. O. Kulik and R. Ellialtioglu (Kluwer Academic, Dordrecht, 2000), p. 259.
6. E. Teller, Z. Phys. **67**, 311 (1931).
7. L. D. Landau, Z. Phys. **64**, 629 (1930).
8. N. B. Brandt, E. N. Bogachek, D. V. Gitsu, *et al.*, Fiz. Nizk. Temp. **8**, 718 (1982) [Sov. J. Low Temp. Phys. **8**, 358 (1982)].
9. L. P. Levy, G. Dolan, J. Dunsmuir, and H. Bouchiat, Phys. Rev. Lett. **64**, 2074 (1990).
10. V. Chandrasekhar, R. A. Webb, M. J. Brady, *et al.*, Phys. Rev. Lett. **67**, 3578 (1991).
11. D. Mailly, C. Chapelier, and A. Benoit, Phys. Rev. Lett. **70**, 2020 (1993).
12. C. Schönenberger, A. Bachtold, C. Strunk, *et al.*, Appl. Phys. A **69**, 283 (1999).
13. B. L. Altshuler, A. G. Aronov, and B. Z. Spivak, Pis'ma Zh. Éksp. Teor. Fiz. **33**, 101 (1981) [JETP Lett. **33**, 94 (1981)].
14. A. Barone, T. Hakioglu, and I. O. Kulik, cond-mat/0203038.
15. I. O. Kulik, T. Hakioglu, and A. Barone, Eur. Phys. J. B **30**, 219 (2002).
16. I. O. Kulik, in *Towards the Controllable Quantum States: Mesoscopic Superconductivity and Spintronics*, Ed. by H. Takayanagi and J. Nitta (World Sci., River Edge, N.J., 2003), p. 302.
17. M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge Univ. Press, Cambridge, 2000).
18. D. Gatteschi, A. Caneschi, L. Pardi, and R. Sessoli, Science **265**, 1054 (1994).
19. K. L. Taft, C. D. Delfs, G. C. Papaefthymiou, *et al.*, J. Am. Chem. Soc. **116**, 823 (1994).
20. H. F. Cheung, E. K. Riedel, and Y. Gefen, Phys. Rev. Lett. **62**, 587 (1989).
21. I. O. Kulik and I. K. Yanson, *The Josephson Effect in Superconductive Tunneling Structures* (Nauka, Moscow, 1970; Israel Program for Scientific Translations, Jerusalem, 1972).
22. A. Barone and G. Paterno, *Physics and Applications of the Josephson Effect* (Wiley, New York, 1982; Mir, Moscow, 1984).
23. A. J. Leggett, in *Chance and Matter*, Ed. by J. Souletier, J. Vannimenus, and R. Stora (Elsevier, Amsterdam, 1996), p. 395.
24. K. K. Likharev, *Dynamics of Josephson Junctions and Circuits* (Gordon and Breach, Amsterdam, 1996).
25. Y. Makhlin, G. Schön, and A. Schnirman, Rev. Mod. Phys. **73**, 357 (2001).
26. D. V. Averin, Fortschr. Phys. **48**, 1055 (2000).
27. H.-F. Cheung, Y. Gefen, E. K. Riedel, and W.-H. Shih, Phys. Rev. B **37**, 6050 (1988).
28. S. Latil, S. Roche, and A. Rubio, Phys. Rev. B **67**, 165420 (2003).
29. I. O. Kulik, A. S. Rozhavskii, and E. N. Bogachek, Pis'ma Zh. Éksp. Teor. Fiz. **47**, 251 (1988) [JETP Lett. **47**, 303 (1988)].
30. M. I. Vischer, B. Rejaei, and G. E. W. Bauer, Europhys. Lett. **36**, 613 (1996).
31. Yu. I. Latyshev, O. Laborde, P. Monceau, and S. Klaumünzer, Phys. Rev. Lett. **78**, 919 (1997).
32. J. Yi, M. Y. Choi, K. Park, and E.-H. Lee, Phys. Rev. Lett. **78**, 3523 (1997).
33. B. Nathanson, O. Entin-Wohlman, and B. Mülschlegel, Phys. Rev. Lett. **80**, 3416 (1998).
34. G. Montambaux, Phys. Rev. Lett. **80**, 3417 (1998).
35. R. E. Peierls, *Quantum Theory of Solids* (Clarendon, Oxford, 1955; Inostrannaya Literatura, Moscow, 1956).
36. H. J. Jahn and E. Teller, Proc. R. Soc. London, Ser. A **161**, 220 (1937).
37. W. A. Harrison, *Electronic Structure and the Properties of Solids* (Cambridge Univ. Press, Cambridge, 1972; Mir, Moscow, 1983).
38. J. M. Ziman, *Principles of the Theory of Solids*, 2nd ed. (Cambridge Univ. Press, London, 1972; Mir, Moscow, 1966).
39. C. Kittel, *Introduction to Solid State Physics*, 7th ed. (Wiley, New York, 1996; Nauka, Moscow, 1978).
40. L. N. Bulaevskii, Usp. Fiz. Nauk **115**, 263 (1975) [Sov. Phys. Usp. **18**, 131 (1975)].
41. I. O. Kulik, in *Technical Proceedings of the 2003 Nanotechnology Conference and Trade Show*, Ed. by M. Laudon and B. Romanowicz (Computational, Boston, 2003), Vol. 2, p. 531.